

Affective Systems: Progressing Emotional Human-Computer Interactivity with Adaptive and Intelligent Game Systems

Maximilian Croissant

Doctor of Philosophy

University of York
Department of Computer Science

September 2024

Dedication

to anyone interested in the topic

Abstract

In principle, affective game systems use the relationship between player emotions and video game content to enhance play motivation, increase engagement and enjoyment, and even facilitate health benefits. However, because of theoretical uncertainties in psychological emotion research and a mix of methodological standards in games research, the best means of creating and evaluating them remains unclear. To advance our understanding of affective game systems, this thesis investigates the emotional player-game feedback loop through multiple pathways. It provides a synthesis of relevant research disciplines (Chapter 2) and a systematic review of current affective game research (Chapter 3) to investigate current theoretical and practical issues in the field. To address these issues, it then presents a framework for developing and evaluating affective game systems (Chapter 4). The framework is evaluated through the development of a new video game and a large-scale randomized controlled comparison study (Chapter 5). Further studies (Chapters 6-7) provide additional validation by making use of the framework to explain emotion measurement and elicitation within specific game contexts. Finally, the future of affective systems is examined, focusing on the role of large language models in overcoming historical barriers. New architectures for language model-driven game agents are proposed, highlighting the potential of this technology in affective computing (Chapter 8). Overall, this thesis proposes new approaches to understanding player emotions and provides standardized and validated methods to develop and evaluate affective games. This thesis aims to shed light on the nature of affective systems how they are currently being developed and evaluated, and how they can be improved to maximise potential benefits.

Contents

Abstract	iii
List of figures	ix
List of tables	xv
Acknowledgements	xix
Declaration	xxi
1 Introduction	1
1.1 Emotions in Video Games	3
1.1.1 Possibilities	5
1.2 Objectives	6
1.3 Structure of the Thesis	7
2 A Tale of Progress and Disagreement	11
2.1 Emotion Research	11
2.1.1 Affective States	12
2.1.2 Theoretical Perspectives	13
2.1.3 Common Contentious Aspects of Emotions	17
2.1.4 Theory Overlap	20
2.1.5 Conclusion	23
2.2 Design and Emotions	24
2.2.1 Design Research	25
2.2.2 Game Design	27
2.3 Affective Computing	30
2.3.1 Affective Gaming	30

2.3.2	Affective Games and Emotion Models	33
2.4	Conclusion	35
3	The Theory-Practice Gap	37
3.1	Aims and Research Questions	37
3.2	Methods	39
3.2.1	Data Collection	39
3.2.2	Search Terms	39
3.2.3	Inclusion Criteria	39
3.2.4	Exclusion Criteria	40
3.2.5	Data Analysis	40
3.2.6	Coding	41
3.3	Results	43
3.3.1	RQ1: Effectiveness of Affective Adaptation	43
3.3.2	RQ2: Emotions in affect-adaptive games	47
3.3.3	RQ3: Methodologies	50
3.4	General Discussion	52
3.4.1	The Effects of Adaptation	53
3.4.2	The role of Emotion	56
3.5	Conclusion	59
4	Narrowing the Gap	61
4.1	Improving the Theory	63
4.1.1	The Player	63
4.1.2	The Game	65
4.1.3	Closing The Loop	65
4.1.4	Summary	66
4.2	Improving the Practice	67
4.2.1	Theory about Target Emotion	68
4.2.2	First domain: Design Emotional Output	70
4.2.3	Second domain: Integrate Emotional Input	73
4.2.4	Summary	77
4.3	Conclusion	79

5	Affect in Adaptive Systems: Creating "The Flow Experience"	81
5.1	Aims and Motivation	82
5.2	The Game: <i>The Flow Experience</i>	84
5.2.1	Gameplay	85
5.2.2	Visual and Sound Design	86
5.2.3	Pilot Study	86
5.3	Study: Testing Phase	86
5.3.1	Materials and Methods	86
5.3.2	Results	90
5.4	Study: Adaptation Phase	94
5.4.1	Materials and Methods	94
5.4.2	Results	96
5.5	General Discussion	97
5.5.1	Associations Between Game Characteristics and Player Emotion	97
5.5.2	Adaptation Strategies Based On Emotional Information	100
5.6	Conclusion	102
6	A Case for Emotional Output	105
6.1	Background	107
6.2	Materials and Methods	109
6.2.1	Dataset and Preprocessing	109
6.2.2	Analysis	112
6.2.3	Ethics Statement	112
6.3	Results	112
6.4	Discussion	116
6.4.1	The Role of Relaxing Games in Increased Daily Player Peaks	117
6.4.2	Means of Relaxation	119
6.5	Conclusion	121
7	A Case for Emotional Input	123
7.1	Background	125
7.2	Materials and Methods	126
7.2.1	The Underwood Project	126
7.2.2	Measures	128

7.2.3	Procedure	129
7.2.4	Participants	130
7.2.5	Analysis	131
7.2.6	Ethics Statement	132
7.3	Results	132
7.4	Discussion	135
7.5	Conclusion	138
8	The Future of Affective Systems	139
8.1	Related Work	140
8.1.1	Affective Agents	140
8.1.2	Language Model Approach	142
8.1.3	Appraisal-Prompting Strategy for Emotion Simulation	144
8.2	Experiment 1: Investigating Situational Emotional Understanding Using Appraisal-Prompting	147
8.2.1	Materials and Methods	147
8.2.2	Results	148
8.3	Experiment 2: Content of an Appraisal-Based Chain-of-Emotion Architecture	149
8.3.1	Materials and Methods	149
8.3.2	Results	152
8.4	Experiment 3: User Evaluation of Game Implementations	153
8.4.1	Materials and Methods	154
8.4.2	Results	157
8.5	Discussion	159
8.5.1	Emotional Intelligence in Language Model Agents	160
8.5.2	User Interaction with Chain-of-Emotion Language Model Agents . .	161
8.5.3	Limitations	162
8.6	Conclusion	164
9	General Discussion and Conclusion	165
9.1	Summary of Empirical Work and Contributions	165
9.1.1	Chapter 3	165
9.1.2	Chapter 4	167
9.1.3	Chapter 5	168

9.1.4	Chapter 6	169
9.1.5	Chapter 7	170
9.1.6	Chapter 8	172
9.2	Progressing Affective Systems	173
9.2.1	Emotion Design for Game Designers	173
9.2.2	Emotion Design for Researchers	174
9.2.3	Emotion Design for Affective System Engineers	175
9.3	Limitations and Future Directions	176
9.3.1	The Emotion Design Framework	176
9.3.2	Understanding Players	178
9.3.3	The Future of Affective Games	180
9.4	Conclusion	181
	Appendix	183
	A Supplementary Material	183
A.1	Chapter 3: The Theory-Practice Gap	183
A.2	Chapter 6: Beyond Adaptive Games	195
A.3	Chapter 7: The Future of Affective Systems	206
	B The Flow Experience: Design Rationale	211
B.1	Narrative	212
B.2	Gameplay	214
B.3	Levels and Progression	216
B.4	Emotion Design Implementation	216
	Abbreviations	219

List of Figures

1.1	Concept of an emotional feedback loop as an illustration of the ongoing affective relationship between a game system and a player, adapted from Sundstrom [27].	3
1.2	Illustration of chapter structure.	8
2.1	Graphical illustration of the circumplex model, adapted from Posner [63] . .	14
2.2	Graphical illustration of the cone model with distinct categories by Plutchik [82]. The vertical dimension represents intensity, and the circle represents similarity among emotion families that are categorized through eight primary emotions: Rage, vigilance, ecstasy, admiration, terror, amazement, grief, and loathing.	16
2.3	Simplified illustration of interactions between multiple emotion components. Most models of emotion components argue for simultaneous interactions between components and as a consequence their influence on each other in multiple directions (see e.g. [98]), as illustrated with the double-sided arrows.	21
2.4	Illustration of the framework of product experience by Desmet and Hekkert [120].	26
2.5	Illustration of the MDE framework by Robson et al. [49]	29
2.6	Illustration of the technical implementation of an affective feedback loop adapted from Bontchev [44]. Physiological data is measured at time t (indicated as $r(t)$ at an optimal set-point level (r_0); $s(t)$ is the adapted game stimuli (e.g. difficulty level) in time t ; $\epsilon(t)$ is the error, i.e. the difference between r_0 and $r(t)$	32
3.1	Flow diagram of data extraction process.	41
3.2	Counts of included studies by year.	44

4.1	Illustration of the emotional feedback loop as an expression of game-player interaction. The game system uses emotional input that infers information about player emotions and adapts the emotional output accordingly. The player appraises the output on multiple variables which leads to interactions on measurable emotion components that can in turn be used to close the feedback loop.	63
4.2	Illustration of the practical framework: To elicit a target emotion, both the domain of the game material and the domain of the user input need to be considered. Potential game characteristics (domain 1) and user input variables (domain 2) that could relate to the target experience need to be identified (step 1), and tested in user research (step 2). The gathered information can be used to introduce adaptivity of the game material based on user input to bring the product closer to the intended experience (step 3) or to improve research by further developing theoretical assumptions about the target emotion (dashed arrows)	68
4.3	Model of relevant variables for the user research that allows the connection of emotional output and emotional input through experience measures given a game context. The tested statistical relationships can be used to inform the design of emotional evocative games, without the need for specified emotion models.	73
4.4	Example of process application using melancholy as an aesthetic experience.	78
5.1	Illustration of the structure of this chapter: Each empirical experiment targets a specific process in the emotion design framework. Study 1 showcases the user research phase and the resulting information is used to build an affect-adaptive game, which is evaluated in Study 2 against a control game.	82
5.2	Illustration of the relevant emotional information to design and test affect-adaptive video games. We describe information that could infer the emotional state of the player as potential emotional input. Information that could influence the emotional state of the player is described as potential emotional output. In order to create a functioning affective loop, the relationship between subjective experience and both potential emotional input and output must be examined.	83

5.3	Screenshots of the first (top left), second (top right and bottom left) and last (bottom right) experimental level of The Flow Experience. Top: The player evades projectile attacks by the enemy through simple movement. Bottom Left: The player blocks a row of incoming projectiles with a shield. Bottom Right: The player dashes through wave attacks to not get hit. The top bar in each screenshot indicates the current state of the level, i.e. the time remaining.	85
5.4	In-game implementation of the Affect Grid adapted from Russel [276]	88
5.5	Bar graphs of the interactions. Left: Effect of attack speed on valence change by level (game context). Right: Effect of attack speed on valence change by expressions of the linear player model consisting of the number of blocks and number of deaths. Error bars symbolize standard error (SE).	93
5.6	Bar graphs of the effects on valence change. Error bars symbolize standard error (SE).	97
6.1	Illustration of the emotional output testing phase of the Emotion Design Framework. The core assumption of the framework is that the statistical relationship between game material and experience is dependent on context and individual differences and needs empirical testing. Assumptions about this relationship might lead to misguided conclusions.	107
6.2	Flow diagram of the number of games screened and included in the study.	111
6.3	Empirical quantile-quantile plot of the covariate. Depicted are the quantiles of the covariate (ADPPs 2019) for the relaxing and control group before and after matching. Small distances between sample points and the diagonal indicate close similarity.	113
6.4	Propensity scores of matched and unmatched samples. Relaxing games are the treated units and control games are the control units.	114
6.5	Scatterplot of the relation between ADPPs in 2019 and ADPPs in 2020 by groups. Scales have been log-transformed for clearer data presentation.	116
6.6	Boxplots of the differences in ADPPs in 2020 by groups. Scales have been log-transformed for clearer data presentation.	117

7.1 Illustration of the emotional input testing phase of the Emotion Design Framework. It is assumed that as with the emotional output, relationships between player input (gameplay, physiology, movement, etc.) and subjective affect are dependent on context and individual differences. 124

7.2 Screenshots of the Underwood VR experience. Top row left to right: Neutral starting room, neutral recovery room, lift to basement level; bottom row left to right: anxiety-inducing starting room, anxiety-inducing room with a lantern, late anxiety-inducing room. 127

7.3 Illustration of the PCA results. The X-axis shows the new components, the Y-axis shows the original predictors. 133

7.4 Scree plot showing the explained variance ratio of the individual principal components. 133

7.5 Illustration of mean subjective valence over time of the sleep-rested group (left) vs. sleep-deprived group (right) for the duration of the Underwood project experience sampled at 1Hz. 134

7.6 Illustration of interaction effects between condition and PC1 as well as condition and PC2 on predicted valence. The valence prediction made by the components changes significantly depending on the group condition (sleep rested vs. sleep deprived). 135

8.1 Schematic representation of the proposed architecture. Agents are set up by providing relevant context and role-playing character information already integrated into a memory system. In interactions, user input gets stored in the memory system and triggers appraisal (i.e., explicit emotion expression) that is also stored in the memory system. Based on the current state of the memory system, agent output is generated. 146

8.2 Example of model input and output for the three conditions. The input of the No-Memory and Memory condition is the same for the first item. For the No-Memory condition, all following items only include the example question and answer, as well as the next question in the scale. The Memory condition includes all prior questions and generated answers. The Appraisal-Prompts condition is the same as the Memory condition, but the example answer is changed to include two steps: First, appraising the situation to generate emotions of the involved person and second, providing the answer. 148

8.3	Results of the comparison between conditions. The Y-axis represents cumulative STEU score by item and the X-axis represents individual STEU items.	150
8.4	Illustration of prompting strategies for all three conditions. The No-Memory condition constructs a prompt out of the system instruction and the user input. The Memory condition constructs a prompt out of the system instruction, message history, and user input. The Chain-of-Emotion condition uses a separate LLM call to appraise the agent’s emotion for each message before creating a user response. It therefore constructs a prompt out of the system instruction, the message history, the history of the emotions generated through the appraisal step, and the user input.	152
8.5	Screenshots of the conversational game “Wunderbar”. The left screenshot shows dialogue provided by the model. The user can click to continue each dialogue line until the input field for a response appears (right screenshot). .	155
A.1	Mean ADPP by time for relaxing games vs control games.	195
B.1	Title screen with logo of the released version of The Flow Experience [5] . .	211
B.2	Level description screen of "Zvaigzne" from the Flow Experience. Each level represents a different symbol and provides some information about the symbolic ideas, historic value, and gameplay connection.	213
B.3	Screenshots of some levels of The Flow Experience. Top Left: Example of unique projectiles to evade through movement. Top Right: The dash is used to escape from non-guardable projectiles. Bottom Left: Example of different cultural ideas communicated through small theming details. Bottom Right: A more intense late-game level with many projectiles to evade.	215
B.4	Screenshots of the level select screen of The Flow Experience. Solving a level (any variant) will open up all outgoing connections to the next levels. .	217
B.5	Screenshots of the "Spirits" screen of The Flow Experience, introducing gameplay customisation based on player preference. This was included as a way to provide gameplay adaptations based on individual preferences. . . .	218

List of Tables

2.1	Affective states with some of the identified design features (see [52], [53], [55]). Event focus refers to the connection of the state to a triggering event (see also affect intention [53]). More potential design features include the synchronization between organismic subsystems (Chapter 2.1.3), behavioural impact, and appraisal involvement [55].	13
2.2	Selection of proponents of discrete emotion models, their collection of fundamental emotions, and the basis for inclusion for these emotions by Ortony et al. [77].	15
2.3	Summary of major foci of emotion theory perspectives, with explanation for emotion elicitation, and differentiating factors. Adapted from Scherer [28], [55]	18
3.1	List of genres for adapted games analyzed in this review.	45
3.2	List of included studies, outcome variables, outcome assessment instrument, effect direction (non-significant effects marked with n.s.), and observed effect size if sufficient information was provided.	45
3.3	List of included studies, definition of underlying emotion structure, emotional state labels, emotion measure, whether the measure was tested in the study, the in-game adapted material used for emotion elicitation, and whether the effect of this material on emotion elicitation was tested in the study.	47
3.4	Included studies, sample size (N), reported demographics (% male and mean age), statistical test, estimated power assuming a medium effect (0.5 SD), target effect size (ES), and risk of bias (RoB; + refers to low risk; +/- refers to some concerns; - refers to high risk of bias).	50
5.1	Descriptive data over all three experimental levels by conditions.	91

5.2	Descriptive data for each experimental level and adaptation condition. . . .	96
6.1	Release years of games in the relaxing group and control group (2008-2013).	113
6.2	Release years of games in the relaxing group and control group (2013-2019).	113
6.3	Genres of games for the relaxing group and the matched control group. . . .	114
6.4	Sample size (N), mean (M) and standard deviation (SD) for the ADPPs in 2019 and 2020 by groups.	115
7.1	List of predictors by type and the corresponding aggregation function for the down-sampling step.	129
7.2	Correlations between each predictor and valence with corresponding p-value.	132
7.3	Results of the mixed linear regression model.	135
8.1	STEU scores (out of 42) by condition. Each STEU item can either be right (1) or wrong (0)	149
8.2	Descriptive overview of LIWC variables per output sentence by condition for the fixed prompt responses with F and p values of the significance test. .	154
8.3	Descriptive overview of user research variables per condition with F and p values of the significance test. Each item has a minimum value of 0 and a maximum value of 6.	158
8.4	Descriptive overview of LIWC variables per participant by condition for all outputs generated in the user study with F and p-value of the significance test.	159
A.1	Complete list of studies included in the review with summaries of aims, methods, and conclusions.	183
A.2	Complete list of games included in the final sample	195
A.3	Instruction, input, and responses for Experiment 2 of Chapter 7	206

Acknowledgements

A defining moment in my academic journey was when I was asked why I would make my project harder than it needed be. There was an expectation of finding novelty that – if any researcher would be so lucky – happened to reflect truth, and not, how I might have possibly assumed, to find truth in what never could be novel. Maybe research is at times like being a cage in search of a bird.

Academic success can mean many things – in this case, I believe it means coming out on the other side with a piece of writing that could confidently be called a “Work in Progress”. But regardless of the meaning, I am sure the single most important predictor of success is social support and I’ve been privileged enough to have experienced that throughout my time as a Ph.D. candidate.

I am very grateful to my supervisors Cade McCall and Guy Schofield. It cannot be taken for granted to have supervisors who genuinely cared - and not only about the work, but also my personal development. Thank you for supporting me through every new idea (the good and bad ones), every change in plan, every roadblock and success.

I thank my advisors David Zendle and Laurissa Tokarchuk. Your interest and ideas helped to shape the thesis to what it now is.

I’m thankful for the IGGI program, my cohort full of talented and interesting people, and the admin team that helped throughout with various resources and opportunities.

I thank my family and my friends for having always inspired (and continue to inspire) my curiosity and creativity.

I’m thankful for you, Madeleine, and incredibly proud for everything you have accomplished during these last few years.

Declaration

I, Maximilian Croissant, declare that this thesis is a presentation of original work and I am the sole author under the supervision of Dr. Cade McCall (primary supervisor) and Dr. Guy Schofield. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Maximilian Croissant wrote the complete original draft of each chapter of this thesis. Cade McCall and Guy Schofield each provided feedback on the drafts of all chapters. Some parts of this thesis have been published in conference proceedings and journals:

- **Chapter 3:** Maximilian Croissant, Guy Schofield, Cade McCall. Theories, methodology, and effects of emotion-adaptive games: A systematic review. In *Entertainment Computing*, Volume 47, August 2023 [1]. All authors conceptualized and designed this study. Maximilian Croissant conducted data search, data extraction, and analyses.
- **Chapter 4:** Maximilian Croissant, Guy Schofield, Cade McCall. Emotion Design for Video Games: A Framework for Affective Interactivity. In *ACM Games: Research and Practice*, 1(3), 1-24 [2]. Maximilian Croissant conceptualized this study, curated relevant literature, and synthesized results. Cade McCall and Guy Schofield provided feedback.
- **Chapter 5:** Maximilian Croissant, Madeleine Frister, Guy Schofield, Cade McCall. The Flow Experience: Evaluating Emotion Design Principles Within a 2D Action Game. *Accepted at International Journal of Human-Computer Interaction..* Maximilian Croissant, Guy Schofield, and Cade McCall conceptualized and designed this study. Maximilian Croissant and Madeleine Frister jointly conceptualized, designed, and developed the video game "The Flow Experienced" used in this study. Maximilian Croissant was responsible for data acquisition and all statistical analyses.

-
- **Chapter 6:** Maximilian Croissant, Madeleine Frister. A data-driven approach for examining the demand for relaxation games on Steam during the COVID-19 pandemic. In *PLoS ONE*, Volume 16, December 2021 [3]. Both authors conceptualized and designed the study. Madeleine Frister investigated relevant literature. Maximilian Croissant curated data and conducted all analyses.
 - **Chapter 7:** Novel analyses of data from a study by Emma Sullivan, Cade McCall, Maximilian Croissant, Lisa-Marie Henderson, Guy Schofield, and Scott Ashley Cairney, “Sleep deprivation influences in-the-moment arousal during exposure to prolonged threats” (preprint; in review). The analyses and research questions in this chapter are all novel. They were conceptualized and designed by Maximilian Croissant, Guy Schofield, and Cade McCall. Statistical analyses were conducted by Maximilian Croissant. Maximilian Croissant, Guy Schofield, and Cade McCall conceptualized and designed this study.
 - **Chapter 8:** Maximilian Croissant, Madeleine Frister, Guy Schofield, Cade McCall. An Appraisal-Based Chain-Of-Emotion Architecture for Affective Language Model Game Agents. In *PLoS ONE*, Volume 19, May 2024 [4]. Maximilian Croissant, Guy Schofield, and Cade McCall conceptualized and designed this study. Maximilian Croissant and Madeleine Frister jointly conceptualized, designed, and developed the video game demo used in this study. Maximilian Croissant curated data and conducted analyses.
 - **Game Project:** Maximilian Croissant, Madeleine Frister. The Flow Experience. [Game]. Vanilla Noir, February 2022 [5].

Copyright © 2024 by Maximilian Croissant

The copyright of this thesis rests with the author. Any quotations from it should be acknowledged appropriately.

Chapter 1

Introduction

"I don't want to be at the mercy of my emotions. I want to use them, to enjoy them, and to dominate them."

— Oscar Wilde, *The Picture of Dorian Gray* [6]

People love to play. The video game industry has experienced steady growth over the years, leading to a market size value of over 160 billion US dollars in 2020 [7], [8] and a forecast of over 290 billion US dollars in 2027. More than 65% of US citizens play video games at least one hour per week [9], similar to the over 51% of players in Europe [10]. There is no question about the ever-growing popularity of video games that has led to significant advancements in video game research over the last years, addressing a variety of questions such as: Why do people play? How can we better understand players and their behaviour by studying games? What are the positive or negative effects of games on health and behaviour? How can we improve video games from a technical, mechanical, and artistic perspective?

These efforts are easily justifiable by the extreme popularity and sales alone, especially considering that the video game industry is comparatively young with early mainstream games only released in the 1970s. The market is constantly evolving with new hardware and software innovations every year, underlining the role of video games as one of the most popular media in the entertainment industry, as well as a promising sector for areas such as health and education [11], as potentially beneficial contributors to cognitive and emotional abilities [12], and as a prime example of applied human-computer interaction (HCI).

Undoubtedly there is a lot to gain from video game research for both commercial and academic interests and in order to add meaningful contributions to the field, it is necessary

to understand why people are increasingly drawn to video games.

Based on classical research on intended actions, Klimmt and Hartmann argue that video games offer distinct affective benefits that motivate video game play and differentiate the medium from other entertainment products, such as movies or music [13]. The first benefit is the concept of effectance, initially introduced by White [14], which can be seen as a strong feeling of satisfaction that arises by the perception of having an impact on the world. As Klimmt and Hartmann argue, the interactive nature of a video game world is inherently designed to feed into this feeling. The second discussed benefit is self-efficacy, a classical psychological concept originally described by Bandura [15]. It may very well be the case that the rewarding feeling of our sense of mastery at least partly accounts for the emotional appeal of video games. Building on similar psychological concepts, more recent experimental research approaches the motivational pull of video games in a perspective of self-determination theory [16]–[18]. Following this theory, both the appeal for consumers and the actual positive affective effects resulting from playing are based on the basic psychological needs for competence, autonomy, and relatedness. Again, it is argued that the interactivity of video games can provide deeply rooted basic satisfactions that we need in order to be happy. In other words, more so than with passive media, we are in control of what is happening, can make independent decisions that have an impact on the experienced outcome, and - for example in the case of multiplayer games - can relate to others and therefore experience meaningful social interactions. Every year, more and more research applies underlying motivational concepts such as self-determination theory, making the affective rewards of games a central focus point in works about design tools and player experience analysis [19]. In fact, such rewards seem not only crucial in explaining player motivations but have also been associated with the act of watching other people play games [20], extending the potential impact even further.

To put it in slightly different terms, the emotional relationship between players and games seems to explain much of the appeal of video games and consequently plays an important role in utilizing their potential benefits. Because of this, the game design process is primarily focused on user experiences, or as Jesse Schell puts it in the *Art of Game Design*:

“ [...] creating them [experiences] is all a game designer really cares about.”

[21]

1.1 Emotions in Video Games

Of course, the importance of affective experiences has been known in the world of design, long before video games became relevant. There are a number of interactional models, for example for "product emotions" [22], building upon highly influential work of researchers like Norman [23], [24], who investigated the affective relationship between products and customers and described it in simple terms in the title of his book: *Emotion and Design. Attractive things work better.* [23]

If we assume that the high commercial and academic interest in a very much growing video game industry is rooted similarly in an emotional relationship between the games and the players (illustrated in Fig 1.1), the natural consequence would be to try to understand every aspect of this relationship. Not only might this lead to the creation of better and more successful games, but it could also further our understanding of human emotion itself and help unlock possibilities of video games promoting targeted emotional support and well-being, as discussed in current reviews about the positive effects gameplay can provide [25], [26].

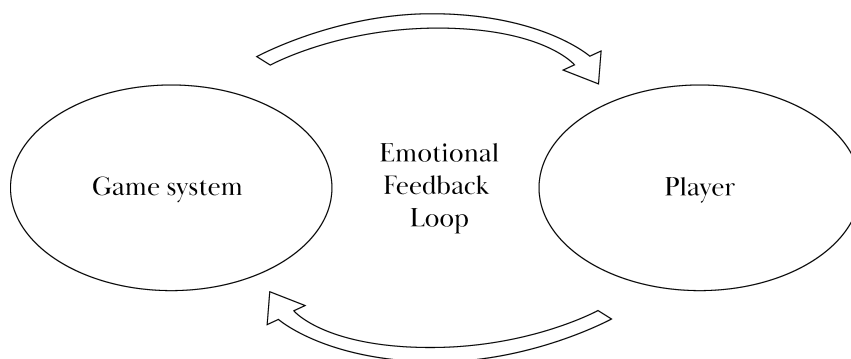


Figure 1.1: Concept of an emotional feedback loop as an illustration of the ongoing affective relationship between a game system and a player, adapted from Sundstrom [27].

Still, understanding this relationship is not an easy task. For a start, emotions as a psychological construct are still a topic of fundamental theoretical debate [28]. Many emotion theoretical perspectives have developed with quite different core assumptions relating to the functions, involved physiological and psychological mechanisms, and expressions of emotions [29]. Not only does this fact still to this day lead to differences in interpreting empirical data depending on the perspective [28], it also creates a gap between the scien-

tific construct of "emotion" and "emotion as part of peoples' everyday life and experience". Everybody has a concept of emotion, everybody knows what anger, sadness, and happiness feel like. Naturally, art is created and consumed with emotions in mind - the connection between art and feeling was studied even in ancient Greece [30]. Still, the precise nature, underlying mechanisms or even basic scientific definitions remain a point of contention among experts [28]. Emotions span countless semantic labels [31], countless theoretical models [29], and countless expressions that are being measured by various instruments [32].

For video games in particular - as both a form of art and an interactional computing medium - this gap between fundamental emotion theories and in practice applied affective technologies becomes obvious. In fact, while we still develop and research theoretical concepts such as "universal and basic emotions" [33], [34], "cognitively appraised emotions" [35], [36], "socially constructed emotions" [37], [38], art-related "aesthetic emotions" [30], "product emotions" [23], and many more, we simultaneously develop and apply technological implementations within video games to make use of these models, for example, to create affect-aware computer agents [39], to measure and interpret emotions in real-time [40] and to adapt the games according to player emotions to optimize user experience [41].

This theory-practice gap has been acknowledged since the early days of "affective computing" research, particularly by the highly influential works of Rosalind Picard [42], [43], who outlined the distinct challenges any affect-aware computer faces, given the theoretical uncertainty around emotions in areas such as emotion sensing, emotion expression, and affect modelling [43]. These include challenges in making reliable recognition efforts in idiosyncratic and poorly defined modes of emotion expressions; the limited availability of accurate cognitive models (and therefore unknown interactions between emotions and other mental components); the technological limitations in sensing, understanding, and expressing emotions; as well as unclear ethical implications when it comes to emotion recognition and targeted elicitation efforts.

Even though this gap is well documented, numerous studies have been conducted to propose models, analyses, and affective game implementations, focused not necessarily on the fundamental challenges in the fields, but rather on the potential benefits of affective games [40], [44], [45]. These benefits are mostly described by three high-level heuristics, originally proposed by Gilleade et al. [46]: "Assist me" as an aspect of affect-aware games to assist players in emotional tasks, "challenge me" as an aspect of affect-aware games

to propose the optimal challenge level for every player, and "emote me" as an aspect for affect-aware games to optimize specific emotional experiences. And while current reviews show some potential of affective games to reach these goals [44], [45], none of the challenges outlined by Picard [43] have been solved, casting doubts on the informative value of some of these empirical efforts. Furthermore, the affective system Gilleade described can be seen as a form of affective adaptation system, i.e. a system that measures, models, and reacts to player emotions [47]. While it is necessary to provide further research to advance complete affect-adaptive systems, there is also considerable interest in understanding emotions for video games that just try to elicit a target experience without real-time adaptation mechanisms. All games try to build an emotional connection with the player [21] - so the most useful version of a scientific understanding of the emotional player-game relationship should not restrict itself to games that adapt its content to player emotions in real time.

With that in mind, the main aim of this thesis is to research and analyse the theoretical issues that currently impact the field of affective systems, and explore potential ways to make meaningful practical contributions to the field, or in other words: to first thoroughly analyze the theory-practice gap and to then narrow it through new methods, theoretical models, and technological architectures revolving around the emotional player-game loop.

1.1.1 Possibilities

The fact that affective video games have the potential to achieve some or all of the goals outlined by Gilleade et al. [46] and expanded by researchers such as Hudlicka [39], [41] is not disputed. It is the general consensus that affective systems in video games could help create better emotional experiences, help different types of players overcome challenges, optimally adapt game material to a target emotion, and even provide health benefits in the form of targeted training or through positive experiences [41]. There is also little doubt about the central role of emotions when it comes to designing and developing games [21], [48], [49] or the importance of deeply rooted psychological mechanisms when explaining play motivation [16], [17].

If we can better understand the underlying concepts grounded in psychology, we will gain a better grasp on the interactive relationship between players and games. This in turn could lead us to build new tools and games that promote the proposed affective gaming benefits. The consequences that would emerge from understanding and narrowing the theory-practice gap could impact future research directions within the fields of psychology,

design, and computer science. Understanding the gap would mean thoroughly analysing work revolving around the current theoretical debate about emotion models, empirical work focusing on affective video games, and other relevant research in design and affective computing. Narrowing the gap would, in turn, mean providing further theoretical ground-work that makes the application of emotion theories in game contexts more robust and certain, and streamlining the process of developing and researching affective games, while not only acknowledging the unique challenges in the field but working towards solutions.

Better understanding the affective player-video game relationship holds the potential to (a) help to overcome some of the challenges connected to applying psychological emotion theories in games; (b) make it easier to understand and develop affective systems; (c) work towards the proposed benefits of affective video games for all games that build an emotional relationship with the player; (c) facilitate the creation of advanced affective artificial intelligent game agents; and (d) assist psychological research in better understanding emotions given a specific interactional context. The theoretical, technological, and empirical contributions within this thesis have the aim to work towards these possibilities.

1.2 Objectives

The main motivation of this thesis is to contribute to the understanding of the emotional interaction between players and games, both in terms of clear, consistent theory and practical tools that allow for the research and development of different affective video game systems. It will integrate contemporary findings from psychology, design, and HCI research into a framework consisting of (a) a robust theoretical model; and (b) a practical guide. Furthermore, this framework will be accompanied by empirical experiments and technical demonstrations of video games to provide data-driven proof-of-concepts both as a way to provide useful guides and as validation efforts for the framework itself. In general terms, this research will be based on one main hypothesis:

A new framework for understanding and developing affective interactivity in video games can be developed that solves theoretical and methodological problems currently present in the field to ultimately guide our understanding of affective games and provide ways to easily develop and test emotional game systems. Ultimately, these efforts can be used to create new system architectures that advance the state-of-the-art in emotional game systems.

In order to properly achieve this aim, the work in this thesis is broken down into multiple sub-objectives as follows:

1. Providing a comprehensive and systematic analysis of the theoretical underpinnings, and methodological approaches of traditional affect-adaptive video games in order to give a clear overview of the potentials and problems in the field that the framework needs to address
2. Developing a framework for affective systems that can be applied to emotion adaptation and emotion simulation with careful consideration of research directions from the perspectives of psychological emotion theory, affective computing theory, and design and affective game theory:
 - (a) The first part gives a theoretical model with clear and useful components to explain and describe affective interactivity
 - (b) The second part gives a practical guide to research and develop affective software while mitigating the risk of making false theoretical assumptions and providing rigorous methodological standards
3. Empirically testing the proposed model and process by developing a new emotion-adaptive video game based on the framework, with a large-scale evaluation of the observed benefits
4. Applying the framework in different contexts beyond full real-time emotion-adaptive games, i.e. games that only implement some affective systems (like elicitation) and not all to help understand the broader implications of the emotional player-game relationship within human-computer interactions
5. Researching new affective systems with progressive new technologies to inform about future possibilities of affective game systems

1.3 Structure of the Thesis

To meet the aforementioned objectives, I have developed a variety of studies and game prototypes with methodologies specifically chosen for their potential knowledge gain for each step. This thesis as a whole utilizes a top-down approach, meaning it was carried

out starting from an overall conceptual perspective to build a solid base for each following contribution. The structure of this thesis mirrors the approach taken:

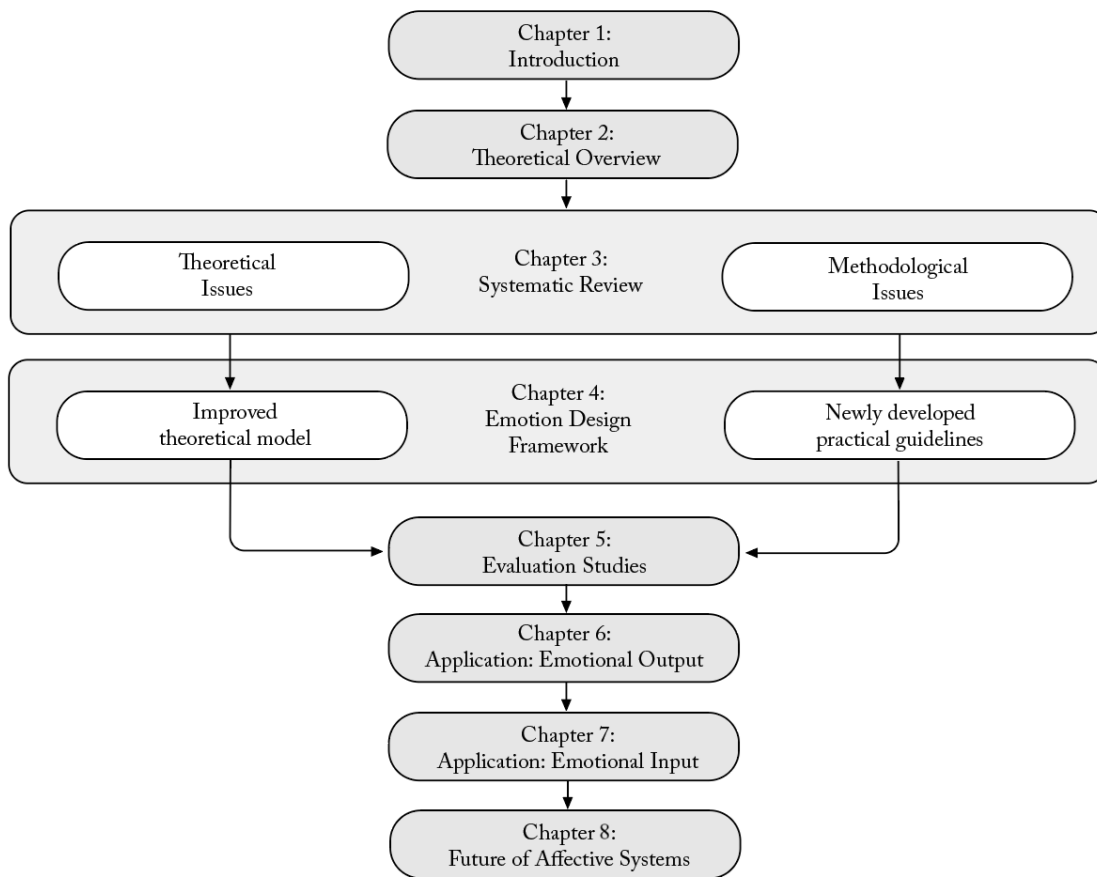


Figure 1.2: Illustration of chapter structure.

Chapter 2: Starting from a general literature overview, this thesis outlines the concept of affective game systems from a psychological, design, and computing perspective. The aim of this step is to provide a deep overview of key literature in each field to better inform the following approaches of the thesis.

Chapter 3: A systematic review analyses more thoroughly a literature sample researching traditional affect adaptation in video games. The aim of this step is to systematically evaluate research in the field in order to better outline the problems regarding theoretical approaches and methodology standards. Results of these chapters are used to inform the development of targeted solutions in both areas.

Chapter 4: This chapter presents the theoretical model and practical guide of the framework. This framework collects and integrates previous findings into a theoretical model based on the affective feedback loop and a practical guide. The framework therefore

specifically targets barriers in theory integration of relevant fields and addresses methodological concerns by providing a flexible step-by-step guide to develop and research affective game systems.

Chapter 5: The newly developed framework is empirically evaluated by applying it to a newly developed affect-adaptive video game. The evaluation is conducted in two steps, mirroring the practical guide of the framework: (a) using the theoretical components of affective interactivity, the statistical relationship between the video game and the player is analyzed; (b) based on the gathered information, an adaptive version is developed and compared to a control version of the game in a randomized controlled trial (RTC) in terms of their effectiveness in eliciting affective positive experiences.

Chapter 6: Focusing more closely on individual components of the proposed framework, this chapter uses concepts and methods of the framework related to emotionally evocative game material presented to the player. It analyzes player behaviour during the COVID-19 pandemic to better understand what makes a game relaxing in times of high stress.

Chapter 7: This chapter uses concepts and methods related to measuring and modelling affective player data to analyze the emotional journeys of participants playing through a virtual reality (VR) horror game to better understand contextual effects on affect prediction and emotion measurement techniques.

Chapter 8: This final research chapter investigates the possibility of using large language models as affective game systems. The aim of this chapter is to test if large language models (known for their abilities in understanding and creating meaningful semantic data) can be used within the emotional game-player loop to function as fully affective systems and what architectural patterns could facilitate the usefulness of such models in affective systems.

Chapter 9: Finally, the general discussion of the thesis evaluates contributions and limitations and provides an outlook for the future of the proposed model, technologies, and the research of affective systems in video games and beyond.

Chapter 2

A Tale of Progress and Disagreement

*"Almost everyone except the psychologist knows
what an emotion is."*

— Paul Thomas Young [50]

In order to better understand the complex relationship between games and emotions, there is a lot of work from multiple disciplines to consider. Naturally, the psychological emotion research perspective provides decades of relevant theoretical groundwork and empirical findings to shape how emotions can be understood and defined. Following a detailed discussion on related psychological work, this chapter will discuss emotions from the perspective of design research, specifically game design research. Finally, the field of affective computing provides us more deeply with models and technological methods that connect affective information from players with computer engines making use of this information in various ways. The aim of the discussion of these bodies of work is to provide an overview of the state-of-the-art regarding emotions in video games that can help to identify knowledge gaps that challenge scientific progress by combining psychological emotion theory with game creation and research [43]. Ultimately, relevant literature from each discipline will be used as a theoretical base to facilitate the creation of an integrative model of emotional player-game interactions.

2.1 Emotion Research

While it seems naturally easy to grasp what the term "emotion" describes, there are a number of issues in precisely defining the term in a scientific context. In 2010, Izard [51] conducted a survey study with 35 highly acclaimed scientists in the field of emotion

research, asking six questions about the definition, functions, and underlying mechanisms of emotions. They found considerable disagreements in almost all answers, with only a 25% agreement in basic definitions of emotions, and even more disagreement in their views of emotion function, emotion elicitation, and the relationships between emotion, cognition, and action. In 2022, most of these disagreements are still not resolved [28].

The fact that emotion as a construct represents multiple different perspectives and interpretations has many important implications. For example, if emotion as a term is not clearly determined, how can it be measured - or to put it differently - what is captured if we attempt to measure it? What exactly can be the promise of emotion-adaptive games if the adaptation is based on a potentially ambiguous concept? How are emotions currently being viewed and applied in games research and what problems emerge through theoretical uncertainties?

Before it is possible to fully unlock the potential of emotions in video games, it seems necessary to first further explore the very foundation of the concept through theoretical and empirical efforts. With this aim in mind, the following sections will discuss the current main theoretical perspectives of emotions with a specific focus on their points of agreement and disagreement. While it is true that there is still no consensus on the exact nature of emotions, there are still decades worth of psychological work researching, refining, and analyzing the phenomenon. A big step in validating the application of emotion concepts in areas like HCI is a thorough understanding of what we currently know about emotions (and clearing up any confusion about these known concepts) and what concepts can be considered to be debated and are therefore potentially subject to change when we gain more knowledge.

2.1.1 Affective States

In many disciplines, the terms "emotion" and "affect" are used interchangeably, which can be a first point of confusion when trying to study emotions. Generally, affect is seen as an umbrella term, describing multiple possible affective states that can be differentiated through certain design features [52]–[54]. A selection of affective states with some of the corresponding design features is collated in Table 2.1.

Emotions are seen as episodic affective states [56]–[58]. This means that they have a fixed duration with varying emotion-specific onset and offset periods of a certain intensity [55]. Differently to states like mood, stances, or attitudes, emotions are characterized by

Affective States	Intensity	Duration	Change Rapidity	Event Focus
Emotion	medium to high	low	high	high
Mood	low to medium	medium	medium	low
Stance	low to medium	low to medium	high	medium
Trait	low	high	absence	absence

Table 2.1: Affective states with some of the identified design features (see [52], [53], [55]). Event focus refers to the connection of the state to a triggering event (see also affect intention [53]). More potential design features include the synchronization between organismic subsystems (Chapter 2.1.3), behavioural impact, and appraisal involvement [55].

a rather short duration, and a high (and rapidly changing) intensity [52]. One additional aspect is the strength of the connection to a triggering event. While affective states such as traits are independent of triggering events, emotions elicitation is based on some form of interaction with an event [52], [54]. While design qualities can be used to differentiate between qualitatively different affective phenomena, it is important that some of these qualities also differ between emotions to a certain extent [55]. However, making clear statements about emotions becomes much easier if each is considered as a specific short and often intense affective experience that can rapidly change and is deeply connected to a triggering event. While states like mood and stance seem valuable as concepts to explore when talking about affect in games, they may potentially be researched independently from emotions due to their different features which influence theoretical assumptions about function, elicitation, and modelling of these states.

2.1.2 Theoretical Perspectives

Even when focusing on emotion as the affective state of interest, questions regarding function, elicitation, and modelling will receive different answers depending on who is asked. There are multiple traditions of viewing emotions in often completely different ways and modern emotion research provides considerable empirical evidence for each perspective. Over the years, the classical theoretical traditions converged in many ways, but there are still highly contradicting aspects in their approaches that cannot be satisfactorily resolved. As a consequence, each perspective will be presented and discussed to provide a comprehensive overview of the state of psychological emotion research.

2.1.2.1 The Dimensional Perspective

Historically, the most important criterion of distinguishing emotions was the pleasantness-unpleasantness dimension [59]. The idea to classify emotions on a scale representing positive and negative emotions developed into one of the most accepted classification criteria and inspired highly-used diagnostic tools, such as the Positive and Negative Affect Scales (PANAS) [60]. Building upon such a unidimensional model, Russell [61], [62] popularized the circumplex model of affect (**Fig 2.1**) which added a dimension for activation or arousal, giving the classification more depth while also providing means of distinguishing between emotions with very similar valence, i.e. with a similar feeling of pleasure.

Some dimensional models also include the dimension of dominance, which is described as a feeling of agency and influence [64]. The pleasure-arousal-dominance (PAD) model of emotion as the basis of highly influential instruments used in emotion application studies, such as the self-assessment manikin (SAM [65]). Recent support for dimensional models of emotion relies upon evidence that patterns of neural activity [66] or peripheral physiological responses [67] map onto such independent emotion dimensions.

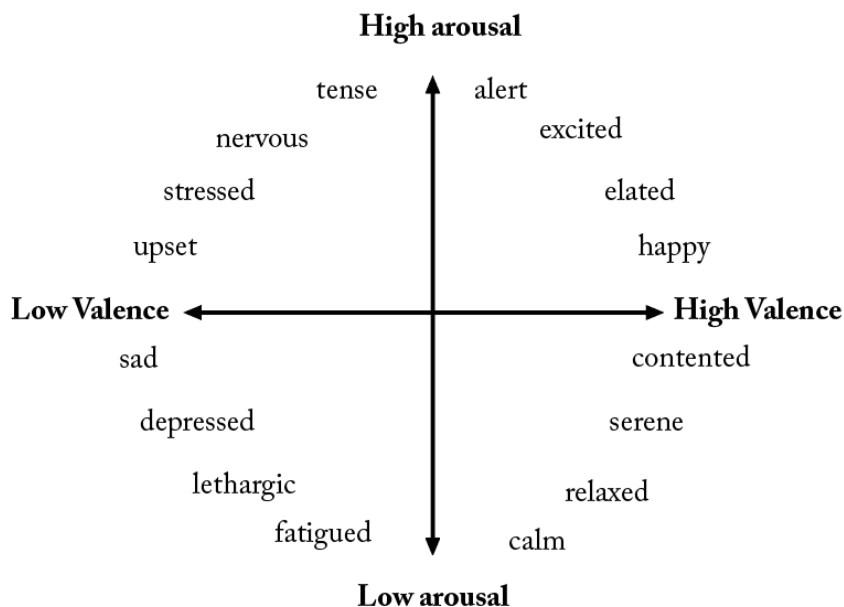


Figure 2.1: Graphical illustration of the circumplex model, adapted from Posner [63]

Reference	Fundamental Emotions	Basis for inclusion
James [68]	Fear, grief, love, rage	Bodily involvement
Watson [69]	Fear, love, rage	Hardwired
Arnold [70]	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness	Relation to action tendencies
Izard [71]	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	Hardwired
Plutchik [72]	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise	Relation to adaptive biological processes
Ekman [73]	Anger, disgust, fear, joy, sadness, surprise	Universal facial expressions
Gray [74]	Rage and terror, anxiety, joy	Hardwired
Panksepp [75]	Expectancy, fear, rage, panic	Hardwired
Tomkins [76]	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	Density of neural firing

Table 2.2: Selection of proponents of discrete emotion models, their collection of fundamental emotions, and the basis for inclusion for these emotions by Ortony et al. [77].

2.1.2.2 The Discrete Perspective

In contrast to dimensional models, discrete emotion models argue for specific and distinct emotion expressions and action motivations. They originate from the pioneering work of Charles Darwin [78], who described lexical emotion terms based on potentially distinct functionality that resulted from evolutionary needs. Modern descriptions of discrete emotions have been popularized by the works of Tomkins [76] and Izard [79], which resulted in the common description of basic emotion categories. Another famous proponent of such a view is Ekman, who provided evidence of universal and therefore culture-independent facial expressions [33]. This has led to the view of emotions as fixed, often biology-dictated constructs. Neuropsychological circuit models mirror this idea by arguing for emotion-specific neural networks for certain emotions, which resulted from an evolutionary necessity [80], [81].

Many proponents of discrete emotion models argue for a collection of fundamental emotions [77]. A selection of some of the most influential research works, their defined fundamental emotions, and their basis for inclusion can be seen in Table 2.2. Today, fundamental emotions are seen as distinct categories or families of emotions and include emotions such as anger, fear, joy, sadness, and disgust. One such model is the cone model by Plutchik (Figure 2.2) which uses eight primary emotions to categorize emotion families.

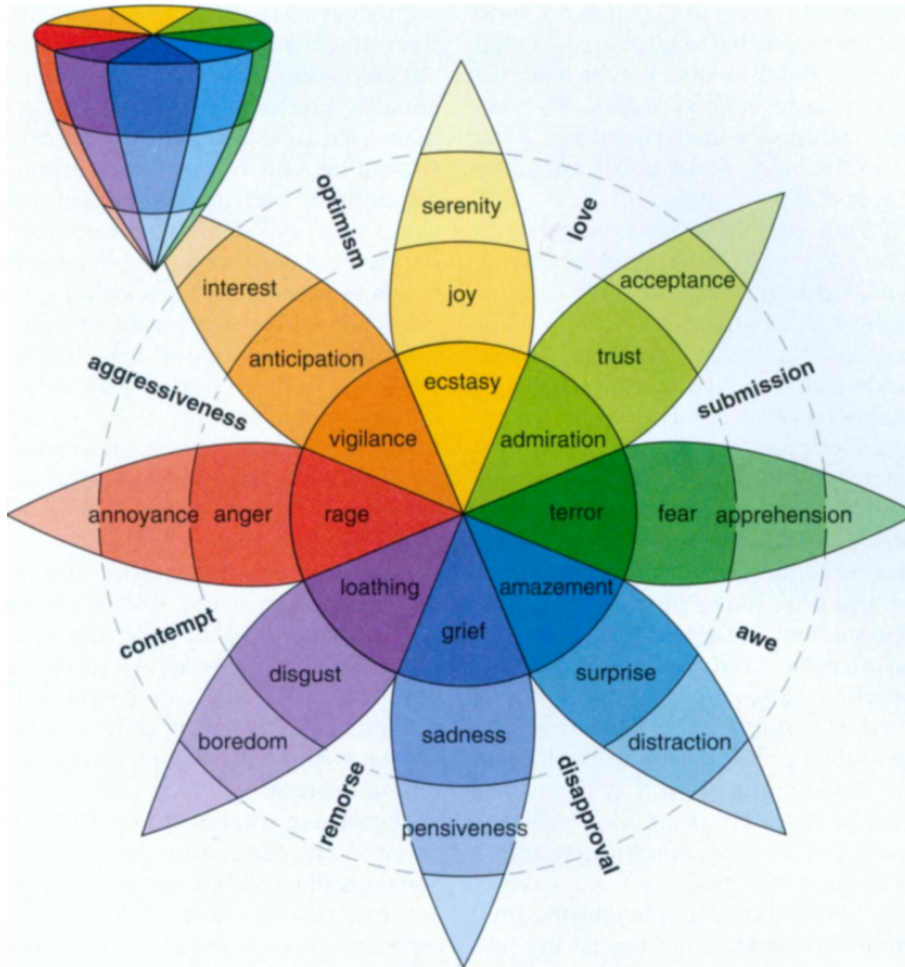


Figure 2.2: Graphical illustration of the cone model with distinct categories by Plutchik [82]. The vertical dimension represents intensity, and the circle represents similarity among emotion families that are categorized through eight primary emotions: Rage, vigilance, ecstasy, admiration, terror, amazement, grief, and loathing.

2.1.2.3 The Constructivist Perspective

The constructivist perspective views emotions as socioculturally determined patterns [83]. A primary factor according to this view is the sociocultural context of situations that elicit emotion through a social interaction. This perspective explains underlying emotional structures resulting from culture-dependent lexical expressions (e.g. [84], [85]).

Contemporary constructivist models of emotion argue that although the “raw materials” of affect are innate, the emergence of a specific emotion is dependent on individual, situational, and culturally specific categorizations [38]. Neurobiological evidence suggests

that specific brain regions do often not correspond to distinct emotional states but rather that multiple more general brain networks are active when experiencing emotions [86], i.e. that emotional experiences are "constructed" through general brain network patterns given a certain context.

2.1.2.4 The Cognitive Perspective

The cognitive perspective puts emphasis on the assumption that emotions are elicited through cognitive evaluations of situations [55]. This idea has been prevalent since ancient philosophers such as Aristotle proposed that differences in emotions stem from differences in beliefs, e.g. one emotion might be labelled anger if it followed a belief of being wrongly accused or shame if it followed a belief of being responsible for something regrettable [87]. In the 1960s researchers like Arnold [70] and Lazarus [88] popularized the concept of subjective and cognitive appraisal of a triggering situation leading to an emotional reaction in modern psychology. According to Lazarus, the appraisal process includes certain appraisal criteria, such as the significance of an event and the believed ability to cope with such an event that relates to specific emotional patterns.

More contemporary research views appraisal as a more "theme"-based approach, meaning that themes in appraisal generate a number of fundamental emotions [89]. Modern appraisal theorists see appraisal as a central emotion component that functions often automatically (i.e. not controlled, rapid, and unconscious), but can also function nonautomatically (i.e. controlled, slow, and conscious) [36]. The process involves the context of a triggering event, as well as individual concerns, history, thoughts, and experiences. The expression of an emotion is determined by how a given situation is appraised which leads to both interindividual and intraindividual differences [90].

2.1.3 Common Contentious Aspects of Emotions

Because different theoretical perspectives argue about emotion mechanisms regarding function, underlying structure, and differentiation, many aspects of emotions are currently not fully understood and our view of these aspects is subject to change with further research. There are a few main points of discussion in modern emotion literature which exemplify the disagreement on fundamental assumptions. It is important to discuss these specific instances to illustrate how examples of applied emotion research could radically differ on a theoretical level, even though the general aim of such applications (e.g. the integration

Perspective	Major Focus	Elicitation Mechanism	Differentiation Mechanism
Dimensional	Underlying dimensional structure	Rarely directly addressed; based on triggering situation	Degree of similarity on dimensions such as valence and arousal
Discrete	Fixed and discrete emotion categories and expressions	Functional reactions based on specific situations	Discrete neuroanatomical circuits or motor programs
Constructivist	Emotions as sociocultural constructions	Cultural interpretation of situations	Socially shares, prototypical representations
Cognitive	Appraisal as a central determining component of emotion expression	Appraisal based on cultural and individual differences	Reaction pattern based on appraisal criteria

Table 2.3: Summary of major foci of emotion theory perspectives, with explanation for emotion elicitation, and differentiating factors. Adapted from Scherer [28], [55]

of emotions in computer programs) might be shared.

2.1.3.1 Underlying Structure

Most notably, regarding the underlying structure of emotions as dimensional or discrete has many implications on how emotions can be differentiated. Most dimensional models agree on the dimensions of valence and arousal (some also include dominance or action tendencies) as primary scales that make emotional states comparable and map them onto a continuum [64]. Discrete emotions on the other hand are seen as distinct categories (such as anger, sadness, etc.) that do not map onto shared dimensions and dimensional descriptions might only be within certain category-specific components (such as the intensity of subjective fear feeling) [91].

The main implication of this uncertainty is the potential problems in describing and measuring emotions. For example, measures of peripheral physiology focusing on the autonomic nervous system (ANS) have been found to inconsistently reflect distinct emotional states in a meta-analysis by Cacioppo et al. [67]. Rather, such measures (for example heart rate monitors) can be used to infer dimensional emotional information, most notably arousal, but also to some degree valence [67]. Behavioural measurements, however, such as facial or body behaviour may convey valence information [32] but has also a significant

specificity for discrete emotional states (see e.g. [92]). Additionally, while emotion terms (and therefore subjective ratings) can be quite intuitively mapped onto one or more dimensions (for example in the circumplex model [63], see Figure 2.1), several findings support the notion of emotion-specific properties, such as the unique involvement of the insula in disgust processing [93]. Quite often, the underlying structure of emotion is assumed based on the possibilities dictated by measurement instruments, making both dimensional and discrete views prevalent and arguably equally important [91]. It is however crucial to acknowledge that neither one nor the other approach can currently be considered as the true underlying structure of emotion and a joined theoretical approach of both perspectives would need clear and universally agreed upon criteria that do not yet exist [94].

As it currently stands, the choice of an instrument often dictates assumptions about the underlying structure of emotions, while it is in truth still unknown. To further strengthen our understanding of emotions, this theoretical deficiency needs a clear acknowledgement, especially when emotion theories and their corresponding assumptions are being applied within other fields.

2.1.3.2 Basic vs. Constructed Emotions

Another point of disagreement is whether emotions represent universal, innate categories or social constructions. Although modern theories agree that both biological and socio-cultural factors play a role in the development and expression of emotions, there are still fundamentally different views regarding the importance and roles of those factors.

Following the logic made famous by Ekman [73], researchers arguing for the existence of basic (or universal) emotions, build their theories on findings supporting across-culture emotion expressions, especially in the face [33], [95], and neurophysiological data examining affective processes related to "old", evolution-shaped systems in the mammalian brain [96], [97]. In this view, emotions are considered hardwired and universal, especially on an unconscious (or "deep") level, while cultural influences begin to play a role on a conscious, second-order level [34], [97].

As illustrated in Chapter 2.1.2.3, the constructivist perspective argues for emotions as sociocultural constructions that do not emerge from innate emotion-specific brain patterns, but that the brain provides mechanisms for affective learning, leading to the construction of emotions within cultural and social contexts [38], [86].

Again, arguments can be made (and are still being made) for both perspectives, al-

though they interpret the nature of emotion completely differently. Basic emotion theories often explain the functions of emotions through an evolutionary lens: Anger and fear lead to approach and avoidance respectively and fulfil therefore different roles in behaviour motivation, dictated by the biological development of humans [81]. Constructivist views on the other hand see learning and sense-making as the main function of emotion, which enables action tendencies, communication, and social influence within experienced interactions [37].

This means that the nature of emotion changes depending on the theoretical perspective. Many questions still need answers before a model can explain all findings related to the emergence and function of emotions and because of that, applications of these models run the risk of inaccurate assumptions.

2.1.3.3 Summary

Because fundamental aspects of emotions are still the subject of debate, it is difficult to make assumptions that are both immediately useful and resilient to changes in our understanding of emotions when applying these concepts to real-life applications. While a given approach might seem expedient for a given design (e.g. dimensional models for biofeedback games), there is a risk of blindly assuming theoretical implications or disregarding important implications of a given theory when modelling emotional reactions. Awareness and acknowledgement of a model's underlying assumptions are necessary for all applications of emotion research in other fields, both to anticipate higher-risk decisions and to accommodate future developments in the field.

2.1.4 Theory Overlap

While many open questions in emotion research remain, years of work focusing on elicitation, structure, and functions of emotions have led to various theoretical assumptions that are more or less shared across emotion models. To enable practical applications in a field with so many theoretical uncertainties, it is important to place special attention on the key overlap between the discussed perspectives. In the best case, these overlaps should be useful in making practical implementations, without relying on contentious assumptions (for a comprehensive analysis regarding theory convergence for psychological research, see [28]).

2.1.4.1 Emotions consist of multiple components

Most theorists acknowledge that emotions have multiple components and that these components influence each other in various ways as depicted in **(Fig 2.3)**. Commonly, emotion components include physiological, behavioural, and subjective (feeling) aspects. Some researchers also include components of motor expressions [98], or emotion regulation [99].

While there is debate about the number and nature of emotion components, it is generally assumed that emotions reflect patterns in componential expressions. For example, both basic emotion theorists and constructivists would argue that physiological, behavioural, and subjective aspects of emotions interact in many ways, although the exact pathway may be unknown or debated.

Many modern theories regard these patterns as key emotion characteristics [35], [100], and some even argue that emotion measurement techniques should capture multiple components for a single emotion to reflect all relevant and distinct aspects [101]. More broadly speaking, there is currently no measure for emotion, or "gold standard" as Mauss et al. describe it [32]. Rather, different measures can map out different componential expressions that can be used to infer specific aspects of emotional states.

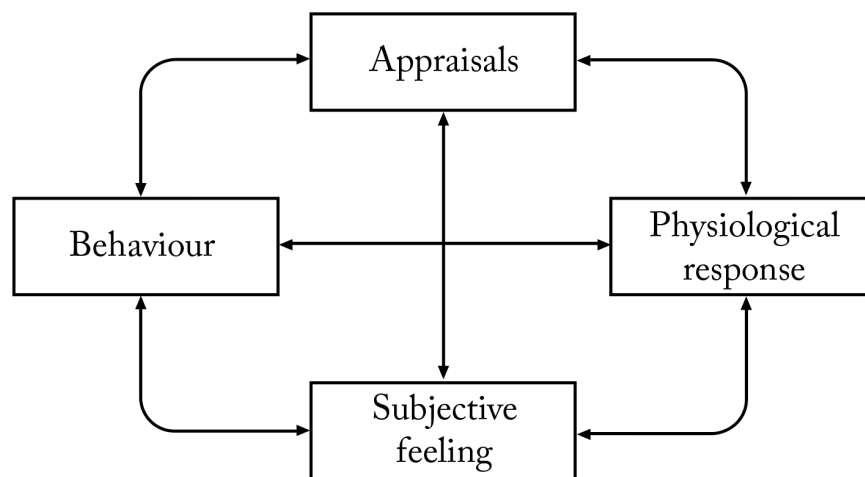


Figure 2.3: Simplified illustration of interactions between multiple emotion components. Most models of emotion components argue for simultaneous interactions between components and as a consequence their influence on each other in multiple directions (see e.g. [98]), as illustrated with the double-sided arrows.

2.1.4.2 Appraisal plays an important role

Emotions result from triggering events, which might be internal (like thoughts or memories) or external (like situations). The notion of emotion appraisal is based on the assumption that emotions are caused by subjective evaluations of such events with regard to their personal significance and the organism's ability to cope [70], [88]. Generally, it is assumed that an event is appraised on multiple variables, including goal relevance, goal congruence, certainty, coping potential, or agency [36]. It is important to note that there are a number of appraisal theories currently under discussion, all involving details that might not be shared. While the fact that appraisal plays an important role in emotion elicitation is currently recognized within all major theoretical perspectives, cognitive theorists claim that the appraisal component is the primary elicitation component that acts on all other components.

While this is not a generally shared assumption, the importance of appraisal is mostly undisputed. The consensus seems to be that multiple pathways of top-down and bottom-up processes interact in the development of emotions [102], which means that appraisal leads to emotional expressions, but is at the same time influenced by other components. Even within cognitive approaches, emotion elicitation is additionally still considered to be highly automatic [90], even though reflective and controlled processes play an important role [103]. This means appraisal includes different processing levels, e.g. automatic sensory-motor processing, automatic perceptual and associative processing, and non-automatic and rule-based conceptual processing [104].

2.1.4.3 Emotions are context-specific

Another shared view is the high emphasis on the role of context in the development of emotions [105]. Following evidence for constructed emotion aspects, most modern researchers emphasize that the context of the emotional material plays an important role in the appraisal and expression of emotions [106]. For example, people sharing a close relationship show similar emotional reactions compared to strangers [107] and culture has been shown to have a big influence on the evaluation of emotional materials and emotions themselves [108]. Furthermore, applied emotion research found that the environmental context of a situation is an important factor in eliciting emotional reactions [109], [110]. For example, how empathy can be elicited with HCI design techniques depends on the material used (e.g. narrative, dialogue, or role-play) and the conceptualization of a user (e.g. as a participant

or observer) [109]. In other words, emotion elicitation and expression are partly dependent on the current state of the environment.

2.1.4.4 Emotions differ between individuals

Besides the influence of situational context, there are individual differences in the appraisal of triggering events and expression of emotions. The same situation can be appraised differently by different people, leading to different emotions [36]. Differences in personality (such as high versus low-risk takers), affective states (such as moods or attitudes), and even biological factors (such as sex) influence emotion processing on a neural and psychological level [111]. Emotions, or emotional expression patterns are highly subjective, regardless of how much the importance of affective learning is emphasized within a given emotion theory. And the notion of a cognitive component (appraisal) provides explanatory power to account for such differences. In other words, emotion elicitation and expression are partly dependent on individual traits.

2.1.4.5 Emotions evolve over time

Emotions are dynamic processes. For example, a threatening situation can evolve from anxious anticipation, to fear of a direct threat, to recovery from that stressor [112]. Some emotional experiences linger while others are fleeting, and the transition between emotions is not always discrete, allowing for the mixing of different emotional states [113]. Important individual differences further affect the degree to which one can “move on” from a given emotional state to another [114], [115]. Still, there is no consensus on the real limits of when an emotion can be considered an emotion and not another affective state. With all this in mind, emotion elicitation and expression are partly dependent on the progression of emotional experiences within a certain time window.

2.1.5 Conclusion

In summary, emotion is a complex term with many uncertainties in its theoretical conceptualization, but also with many well-developed characteristics that are shared between even the most different philosophies. As expected from such a complex psychological phenomenon, it is not easy to define or grasp what an emotion is. As Griffiths [116] puts it:

The proper response to current knowledge is that there is no object of scientific knowledge which corresponds to "emotion".

There is no doubt that emotions do exist and that they can be described in their structure, function, differentiating mechanisms, or in their expressions. But we are not at a point where such a description could be seen as a scientific standard. Emotion currently means many things both in our everyday lives and in the scientific discourse. Misuse or misconceptions within this area seem almost impossible to avoid.

Still, progress is being made and will be made in the years to follow in an effort to further our understanding of the human mind. Processes that prevent both misconceptions and paralysis in light of these theoretical barriers become especially important in applied emotion research.

In the case of player-game interactions, building a theoretical basis on shared assumptions in emotion perspectives provides the possibility to model emotions without limiting the approach to one highly specific theory. Such a model would provide meaningful theoretical groundwork while staying robust to changes in our understanding of emotions. Framing the game-player relationship in such a way would allow developers to focus not only on established theoretical assumptions, but also on specific perspectives with all their implications. Either way, such an approach could reduce confusion that arises through inconsistent or poorly justified applications of theoretical assumptions.

2.2 Design and Emotions

Emotion research is not only relevant for explaining psychological phenomena but also for designing and building products that build an emotional relationship. This includes both the research about how individuals feel or experience certain products and the research about how products evoke certain emotions within individuals. In order to provide more insights into the design of emotions in games, both of these questions have interesting implications. It is therefore necessary to discuss the progress in emotion design (both in general design and in game design) more deeply to gain insights into the affective relationship between a product and its users.

2.2.1 Design Research

For a long time, design research has been concerned with the emotional relationship between a product and a user and has introduced many frameworks to model this relationship. Beginning in the 1990s, the emotional impact of designs gained importance, contrasting a purely functional view on good design. There was high interest, both in the industry (such as the technology industry) and in customers, to make products more engaging, more authentic, and easier to use by building designs based on key consumer emotions [117]. In the following years, Desmet [118] proposed a product emotions model that contained 14 distinct affective reactions and used this to examine how products could elicit emotions. They argued that there is no convincing relationship between a product and an emotion, but rather a relationship between a product and the emotional reaction of a specific person in a specific context. This means that while the process of eliciting emotions is universal, the actual affective response depends on the appraised significance, which can be complex, personal, and dependent on contextual factors like time or environment. The idea that products can be designed from the perspective of being an experience for users established itself as a key school of thought in areas like HCI [119].

Following such a view that puts high emphasis on contextual aspects of products, many frameworks of user-product relationships on an emotional level emerged. One framework in this area was proposed by Norman [24] who conceptualized three levels of design: the visceral level, behaviour level, and reflective level. He argued that designs would affect responses based on these different levels of information processing; i.e. intuitive responses and first impressions on the visceral level, consumption actions on the behaviour level, and reflection of experiences on the reflective level. For Norman, the psychological mechanisms in evaluating a product played a central role in making informed design decisions.

Similarly, the framework of product experience (Fig 2.4) describes three levels of experience: aesthetic experience (i.e. the sensual pleasure resulting in a product interaction), attribution of meaning (i.e. the cognitive evaluation of a product which depends in individual and cultural differences), and the emotional experience (i.e. the elicited emotions based on appraisal, following the first two levels). This framework aimed to provide useful patterns for designers for each of these levels to find more control in product design for emotion elicitation [120]. While also acknowledging underlying psychological mechanisms within the three levels, this framework emphasized the interaction between a product and user, rather than the levels of processing within a user.

The process of designing to elicit an emotional response became known as Emotional Design and was communicated via channels like style, function, form, and usability based on users' needs and demands [121]. A key aspect of this process is user-centred design, meaning a clear understanding of the target user requirements and the context of use for a product [122], mirroring the importance of individual, sociocultural, and situational differences in emotion elicitation. Methods like interviews, questionnaires, focus groups, observations, and simulations were recommended for both understanding users at the beginning of a design cycle and evaluating the design at the final stages [122].

These principles were applied by different studies to showcase how a design could benefit from focusing on emotional value. For example, Desmet et al. [22] showed that a "wow-effect" could be elicited in the design of a cell phone in three steps: First, a usable emotional conceptualization of the wow effect, then the combination of qualitative and quantitative research to provide rich knowledge of products that elicit similar experiences, and third, the focus on the product's character instead of the features to form a deeper relationship. This process showcased an application of how design work could be achieved with a focus on product experiences. Similarly, there are various examples where emotional design was applied to be appealing first, functional second, and finally create a relationship with the user for higher emotional impact [123], [124], often mirroring the

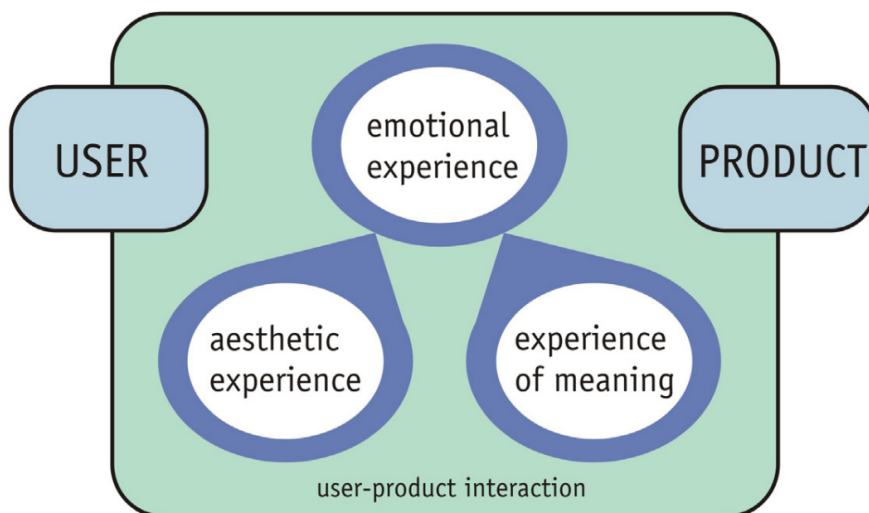


Figure 2.4: Illustration of the framework of product experience by Desmet and Hekkert [120].

three design levels of Norman [24]. Many studies provided evidence for the utility of such approaches, especially regarding the importance of aesthetic appeal and reflective meaning. For example, Sonderegger and Sauer [125] found that aesthetic appeal had an influence on usability ratings in a phone application; and Olsson and Salo [126] provided an extensive qualitative analysis of meaning-related concepts in Augmented Reality (AR) and mobile applications.

To summarize, product experience design puts an emphasis on the relationship between a product and a user and tries to evoke emotions through the design of the product's characteristics that in turn influence psychological mechanisms on different levels of information processing, reaching from highly automatic, fast, and unconscious to controlled and reflective. An important part of this theoretical model is that this relationship is dependent on the user, their expectations, preferences, personality, and even their current affective state [117], [122], which further implies the relevance of circumstance when analyzing this relationship.

2.2.2 Game Design

Naturally, the interest in emotions when designing products has been represented in the game industry since the very beginning, as games were considered entertainment products with the main purpose being the creation of positive experiences [127]. Salen and Zimmerman's influential book *Rules of Play: Game Design Fundamentals* [128] discusses various game design concepts (such as rules, play, and culture) that are used to address second-order design problems, i.e. the indirect design of player experiences by directly designing game rules and environments. More precisely, game experience is seen as a direct consequence of player interactions with game materials and through these interactions, experiences are always evoked [128].

Adding to the notion that PX is the focal point of game design, Schell's *The Art of Game Design* [21] describes strategies that are based on their impact on players, arguing that the whole process of game design should be built around the elicitation of target experiences. This notion has been established in the game industry and provided a working framework for big studios to develop games in a large team by deliberately making decisions for the design of experiences [129]. As a consequence, many tools have been developed to describe techniques for emotional PX design, for example, Freeman's *Emotioneering* [130] that provides methods to design specific game elements, like Non-Player Characters (NPCs),

dialogues, worlds, or plots.

In line with the concepts of PX design, Isbister [131] described game elements that can be specifically used to evoke certain emotions. She argues that games have two unique qualities that are not found in other media and should be the focus point of emotional design in games: The first quality is choice that brings the player close to the action of a game and holds emotional potential that can be supported by the use of avatars, NPCs, customization, social play, physical movements, and communities. Choice is often used in relation to player agency or processes that make an impact in the game world that can result in meaningful emotional reactions [132]. The second quality is flow, originally described by Csikszentmihalyi [133], describing a state of concentration, and deep, immersive enjoyment. Chen [134] argued that each game can evoke an optimal flow state by balancing challenge and player abilities, which would maximise engagement and enjoyment for each game.

These models often very closely focus on positive game experience, "enjoyment" and "fun", but have been criticized for the often inexact or inconsistent relationships between each other and with psychological constructs relating to enjoyment and fun [135]. The player experience of need satisfaction (PENS) [17], [136] model chooses a different approach by utilizing a more motivation-oriented explanation grounded in self-determination theory (SDT) [137]. It explains game enjoyment through intrinsic motivation that satisfies basic psychological needs: Autonomy (the experience of acting through own choice in congruence with own goals); competence (the experience of achieving desired change); and relatedness (the experience of connection with others). While SDT has a lot to offer in terms of explaining player motivation and enjoyment, it cannot account for the full emotional relationship that makes up player experiences.

Consequentially, recent efforts have been made to integrate existing knowledge about the design of emotional reactions from product experience research and make them usable in a game design context. For example, some studies provide adaptations of Norman's [24] three design levels specifically for game design [138], [139], arguing that games should make specific emotion elicitation decisions on a visceral, behavioural, and reflective level. The Mechanics, Dynamics, and Aesthetics (MDA) framework [140] is a popular game design framework that focuses on aesthetic game content as a central design component to elicit emotions and proposes an iterative approach to design for the aesthetic aspects of games. It also provides taxonomies of game elements that are potentially related to emotional reactions (such as narrative and challenge) and models how they relate to game

systems. Similarly, the Mechanics, Dynamics, Emotions framework (MDE; see Figure 2.5) [49] builds on the MDA to propose a model of gamification design. The emotion component here describes a player's state of mind and is used to explain the importance of emotions for a game experience in relationship to game dynamics and mechanics. While these frameworks integrate the concept of emotions as a crucial design component within a game, they are not concerned with the psychological nature of emotions or the affective interaction between player and game, meaning that they put emphasis on the value of emotion in game design, but do not integrate underlying psychological mechanisms in their frameworks.

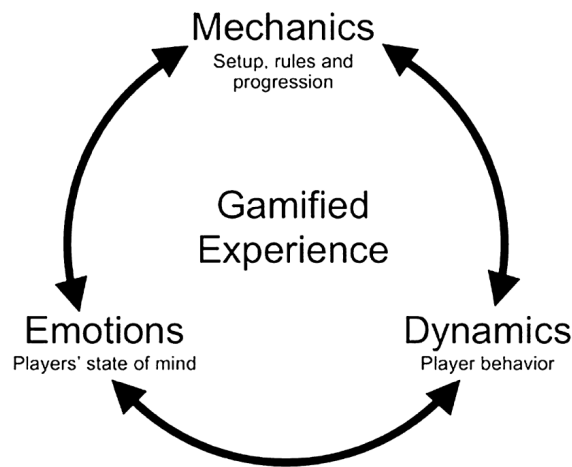


Figure 2.5: Illustration of the MDE framework by Robson et al. [49]

Beyond theoretical models of game experiences, there are many practical frameworks providing methodological tools to help shape game experience design. Deterding [135] reviewed both academic and industry methods for gameful design and presented explicit ways to design for target experiences, such as design lenses and skill atoms. Pichlmair and Johansen [141] conducted an analysis of game design practices that relate to the "feel" of a game, which draws from research concerning the elicitation of specific affective reactions. They identified three main strategies to shape game feel that are currently used: Tuning the physicality of a game; juicing as the act of amplifying game moments; and streamlining support mechanisms to enable players to realise their intentions.

While there are many frameworks modelling the close relationship between emotion and game design, there are still many open questions, specifically about the integration of and application of psychological conceptualizations of emotion components (especially beyond

motivation and enjoyment). Questions include for example how game-related emotions occur, how they can be measured, and how this information can be used to assess players' emotional reactions to a game, as well as how such information can be used to make more targeted affective design decisions or inform emotion theories. Some answers to these questions have been attempted in the discipline of computer science, specifically in affective computing.

2.3 Affective Computing

As a discipline, affective computing has been a prominent topic within HCI research and explored the measurement of and reaction to user emotions by a computer system [42]. Affective computing is therefore concerned with the interactional relationship between technology and users that has the potential to provide optimal emotional experiences by taking the current affective state of the user into account. In an early work on the topic, Rosalind Picard discussed affective computing and its potential to address conceptual uncertainties present in emotion research by focusing on pattern recognition (either physiological or behavioural) and describing emotions in a reasonable manner, without making assumptions about underlying structures [42]. In the following years, much progress has been made that brought the discipline of affective computing into gaming.

2.3.1 Affective Gaming

In an effort to bring affective computing research to games, Hudlicka [41] described principles and current issues of the three main components of affective games: Emotion sensing and recognition, computational models of emotion, and emotion expression or adaptation. Building on this, she outlined requirements for an ideal emotion engine that could accurately measure and interpret emotional data from the player and feed it into a model, as well as create realistic emotional behaviours for NPCs [142]. Affective design in general is therefore mainly concerned with addressing these requirements and developing solutions within three affective tasks:

1. Emotion Sensing: Lux et al. [143] identified 76 studies that use biofeedback devices as an affective measurement, ranging from measures of cardiovascular activity to electrodermal activity, body movement, or respiration. For games specifically, common measurements include physical measures like body movement; physiological measures

like skin conductance, heart rate, muscle movement, or brain waves; and observation measures like facial or vocal expression [144]. Currently, there is no universally accurate instrument to measure emotions and recognition methods depend on emotion model assumptions, individual differences, and context. Furthermore, measurements are often seen as invasive, expensive, and unpractical [145].

2. Computational Emotion Modelling: Models of emotions are most commonly researched in artificial intelligence game studies with the main aim being the development of realistic affective game agents. In a recent review, Hamdy and King [146] collected requirements to develop emotional agents and provided an overview of computational emotion models. They pointed out that models often have to simplify the complex nature of emotions and are also quite costly and difficult to develop. Similarly, Hudlicka [39] found that models often do not address detailed implications of psychological theories. They concluded that in order to fit with modern, complex theories of emotion, believable and realistic agents need to address theoretical uncertainties first, meaning that more systematic and integrative research is necessary. In a systematic review by Wang et al. [147], current practices in emotion modelling for affective computing and their implications were described, uncovering methodological difficulties present in the field.
3. Adaptation: Finally, research considering emotion adaptation focuses either on affect-based changes in agents or the game world [47]. Agents are again used to express emotions based on the underlying model and showcase mostly "believable" emotional behaviour, while the game world is specifically designed to reinforce a target emotion. For example, adaptive difficulty has been used to limit frustration [148], and adaptive camera movement has been used to augment a game's narrative [149].

In order to facilitate research addressing these tasks, Yannakakis and Paiva [47] provided descriptions of three game system modules: an emotion detection module (a module to measure and model player emotions), an adaptation module (a module to adapt the game world to these player emotions), and an elicitation module (a module to elicit target player emotions). These modules are embedded into a shared high-level concept of the emotional interaction between player and game, known as the affective feedback loop [27], [44], [47]. The closed nature of the loop is emphasized, as the ongoing adaptation of the game system to the changes in players' emotions is argued to be a unique characteristic of

games compared to other mediums [150] and is also generally believed to facilitate emotional benefits, such as health benefits, more accessible games, new gameplay opportunities, and higher enjoyment.

Figure 2.6 shows the technical implementation of an affective feedback loop as presented by Bontchev [44] in an example of a biofeedback game: Physiological data is measured via a sensor, detected and then interpreted via negative feedback control. The measured data is compared to optimal data and the difference is seen as an error. The game then adapts to reduce this error by changing the game stimuli (in this example the game’s difficulty) and so dynamically adapts the game based on the player’s emotions. In turn, the new game stimuli evoke an emotional reaction that is continuously measured in a closed loop.

Using such an approach, a number of affective games have been developed and researched, often providing evidence for the benefits of emotion-adaptive games [40]. These benefits fall into four main categories: (a) both mental and physical health benefits as a result of new adaptive training opportunities [151]–[155]; (b) more accessible games with similar experiences for every player [148]; (c) new gameplay opportunities, like body-response control games or affective user interfaces [156], [157]; and finally (d) more emotionally impactful and immersive games [142], [158].

However, because of the current technical limitations, theoretical uncertainties, and poor representation in commercial games, there are problems to solve in the field of affective games before these benefits can be considered applicable for wider use. Not only are

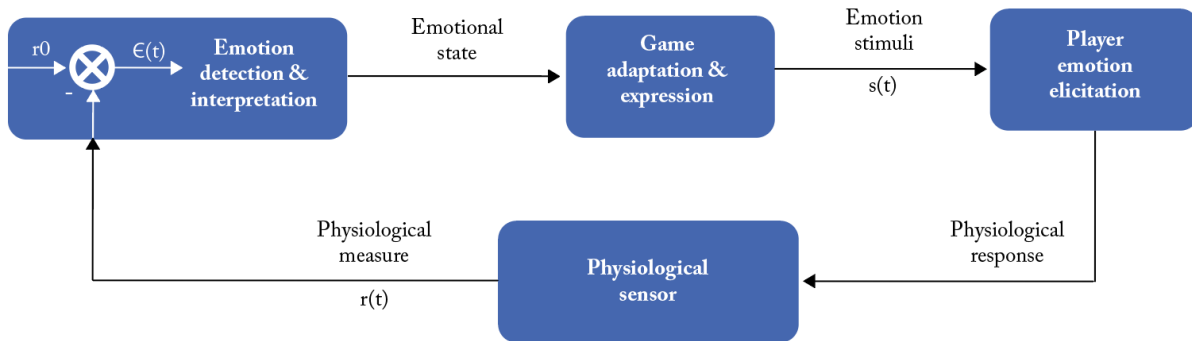


Figure 2.6: Illustration of the technical implementation of an affective feedback loop adapted from Bontchev [44]. Physiological data is measured at time t (indicated as $r(t)$) at an optimal set-point level ($r0$); $s(t)$ is the adapted game stimuli (e.g. difficulty level) in time t ; $\epsilon(t)$ is the error, i.e. the difference between $r0$ and $r(t)$.

conceptual uncertainties present regarding emotion research, but the nature of affective games and how they can be developed is still a field that needs further attention. For example, affective games were originally seen only as biofeedback games [159] and are often still associated with that history, although later definitions focus more generally on adaptivity to unconscious or conscious emotional data [41], [148]. Such definitions could arguably include games that are adaptive in certain areas, such as open-world games like Red Dead Redemption II [160] or Cyberpunk 77 [161] that change their narrative and game world depending on in-game behaviour or, to a limited extent all games with multiple difficulty settings. Nevertheless, affective games are mostly presented in the literature as a certain type of game that measures and reacts to primary emotional data [44]. This conceptualization limits game design decisions that might tap the aforementioned benefits of affective games. Adaptivity to primary emotional data is limited by various theoretical and technological barriers that might discourage design interventions when it comes to the broader idea of affective games. There is therefore a potential to further examine the interactional relationship between game and player that is represented in affective games in a more universally applicable way.

2.3.2 Affective Games and Emotion Models

In a review from 2016, Brontchev [44] analyzed 14 video games that integrated affect-based adaptation techniques. They found that affective-adaptive games generally were effective in achieving goal-oriented changes (e.g. more enjoyment while playing). However, because of often incomplete internal models of affective player behaviour, they conclude that there is much more work to do to achieve a complete and realistic system for affect adaptation in video games. In a systematic review from 2020 Robinson et al. [40] analyzed 162 biofeedback game studies and found effects not only for player engagement but also for treatment in health-related affective games. However, they also note that many affective game studies show insufficient critical reflection, both in terms of how technological limits are reported and how rigorous evaluation is executed.

In addition to these problems, there is a limited body of work inspecting the relationships between affective games and psychological emotion research. This topic has however a big influence on how affective games can be classified within a scientific context. Depending on the methodology, affective games often make assumptions about emotion models that can potentially be very specific and include debated aspects.

For example, to make use of biofeedback data, emotions are most commonly considered in a dimensional perspective and measured in their expression of valence or arousal in games using physiological data [162]. In a review of psychophysiological methods in game research, Kivikangas et al. [163] specified the dimensional model as a standard for affective biofeedback games. Similarly, Robinson et al. [40] scoped their systematic review of affective games to the use of a dimensional emotion model as physiological measures are most commonly restricted to valence and arousal dimensions. In fact, they defined "affect" as the connection between purely physiological states (such as physiological arousal) and mental states (such as relaxation). In other words, affect is considered as the mapping between physiological and mental states based on Russel's circumplex model [61], so by definition existing only in a dimensional space that can be mapped to emotion labels, which is heavily criticized by many emotion researchers [32].

The predominance of dimensional emotion models in game research can be seen as a consequence of the often physiologically based methodology and the aim to make affective data easy to interpret. While most studies in this area report enhanced player experiences by utilizing a dimensional affective game system [40], they do not represent the full range of emotional phenomena and many modern researchers highlight the risk of making strict inferences from arousal or valence values to emotional states [91], [164]. Potentially even more problematic are models that map dimensional values onto distinct emotional states (e.g. interpreting low valence and high arousal as the distinct state of fear) as such an approach is contested by many psychological theorists (see Chapter 2.1.2 for more information).

But there are also game studies applying different emotion-theoretical perspectives. Affective games using cameras to capture facial expressions for gameplay control often base their classification on basic emotion models (e.g. [157], [165]), often building upon the theoretical work of Ekman [33] associating facial expressions with universal emotions. The emotional outputs are used both as distinct states (e.g. [157]), or mapped onto dimensional models to make inferences easier to interpret and combine face recognition techniques with other physiological measures [166], [167].

Following a more constructivist approach, interactional affective computing models, like the one from Boehner et al. [168], propose methods to interpret emotions based on the situational context within human-computer interaction and not based on their physiological expression. This view was adopted by many User Experience (UX) focused studies [169],

[170] and studies on emotional agents within video games [171], [172] that model, interpret, and evaluate emotions from a purely interactional perspective.

Finally, game research that is focused on modelling emotional NPC behaviour, such as GAMYGDALA [173] or the works of Hudlicka et al. [39] implement cognitive models of emotions by introducing appraisal-based mechanisms into the design of game agents. While this is an area that sees more and more contributions, there is again often a mapping from appraisal patterns to distinct emotions (e.g. [173]). While deterministic mapping from appraisal to emotions could be an efficient way to model agents, it does not reflect the complex pathways of top-down and bottom-up processes in human emotions [102].

2.4 Conclusion

This chapter illustrated and discussed key literature in the disciplines of psychological emotion research, design, and affective computing to provide a theoretical background for how the relationship between emotions and games is currently being modelled and applied. There are many research directions working towards our understanding of emotional states, our ability to design emotional experiences, and possible methodological solutions to model emotions and adapt systems to them. These directions promise to facilitate great benefits for both research and design with further contributions to our understanding of the emotional relationship between players and games.

Still, there are many problems in each discipline, ranging from theoretical uncertainties to technical limitations. It is not yet clear what emotions are and how they can be measured and modelled within computer systems to yield the best results. Because research on theoretical models and on applied solutions are being developed continuously, it is also often not clear in what ways these two directions overlap or contradict each other. Emotions and games are complex constructs - and only an awareness of these complexities can lead to a better understanding of their interactions.

To maximize the potential contribution in the field of affective video games, there is therefore a need for clear standards, both in theoretical grounding, as well as in methodological rigour. Building on such efforts, it might be possible to provide improved technical solutions backed with strong empirical data and therefore ultimately better systems utilizing all aspects of the emotional video game-player interaction. In order to achieve this, we need to truly understand affective games: How do they measure and model emotions? What benefits can we expect? And how are they affected by the current state of theoretical

and practical barriers in all their connected fields? A systematic approach to answer these questions could lead to directed solutions and ultimately to better applications of affective human-computer interactions.

Chapter 3

The Theory-Practice Gap

"What does his lucid explanation amount to but this, that in theory there is no difference between theory and practice, while in practice there is?"

— Benjamin Brewster [174]

As Chapter 2 demonstrates, there is a great deal of research on emotion, on the affective relationship between players and games, and on methods for adaption that make adaptive video games possible. Proposed benefits include the potential to enhance player enjoyment, learning or mental health benefits, and new ways to experience games. Still, there are many potential issues that need investigating, especially in terms of theoretical models and methodological approaches. It is not yet clear how well emotion-adaptivity facilitates the proposed benefits compared to other types of games and what methodological approaches could be used to create adaptive games. Therefore, a systematic review is presented here that investigates the current state of affect-adaptive games research in order to uncover the biggest barriers in the field.

3.1 Aims and Research Questions

Prior reviews have found promising effects for emotion-adaptive games [44], [45] and reported an extensive overview of findings and methods specifically for physiology-based games [40]. A similar understanding of the reported effects of emotion-adaptive games would be very valuable in assessing the benefits and risks involved in the design and development process. It is currently not clear how emotion-adaptive games perform against control conditions, what outcomes (such as health benefits or player enjoyment) are being

investigated, and how large reported effect sizes are. Furthermore, it is not clear in which ways affective games are grounded in psychological theories, given the theoretical disagreement in the field. If the true nature of emotions is still not fully explored, how are emotions structurally represented in games and how are they being manipulated? The current practice in developing and researching affective games needs to be tested given the potential theoretical and practical issues, outlined in Chapter 2. How affect-adaptive games address these issues and base their mechanisms on emotional theories is a valuable question in assessing their impact on psychophysiological benefits. Additionally, the quality of provided evidence for the effect of emotion-adaptive games in terms of their methodological approaches is not yet clear and may further provide important data to evaluate the true potential of affective games.

To my knowledge, this is the first systematic review that analyses affect-adaptive video game studies in terms of (a) the effect of adaptation; (b) the theoretical assumptions regarding emotions; and (c) the quality of the evidence regarding evaluation of such studies. This study tries to address these gaps by systematically analyzing the available research body of affect-adaptive video games to answer the following questions:

1. **RQ1: What evidence is there for the effectiveness of game adaptation to player emotions?**
 - (a) How many studies evaluate the effect of affect-adaptation within a video game?
 - (b) What dependent variable is used to indicate adaptation success?
 - (c) What empirical evidence is reported as part of the evaluation?

2. **RQ2: What emotion theoretical assumptions are being applied to build affective adaptation?**
 - (a) How are target emotions defined? What theories are used?
 - (b) What measures are used to indicate affective states and how are they tested?
 - (c) What material is used to elicit emotions and how are they tested?

3. **RQ3: How are affect-adaptive games being evaluated?**
 - (a) What sample characteristics are provided?
 - (b) What control condition is used for the evaluation?
 - (c) What are the characteristics of the methodology?

3.2 Methods

This review follows the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA [175]). All studies that empirically evaluated an affect-adaptive video game by comparing the affect-adaptive game to a control condition were considered for inclusion. A protocol for the study was preregistered on the Open Science Framework (OSF [176]) before data screening commenced, but after the initial database searches, which were conducted first to assess the scope and feasibility of the study.

3.2.1 Data Collection

Electronic databases were searched on April 8th 2022. Databases that are relevant to information technology, health, and social sciences were chosen, which include: ACM Digital Library (n = 561), IEEE Explore (n = 824), Science Direct (n = 53), and Scopus (n = 2490). Additional studies (n = 2) were identified through reference lists of relevant studies [40], [44], as well as through searches on Google Scholar. The database searches returned a total of 3,930 papers.

3.2.2 Search Terms

Search terms were chosen based on three necessary study characteristics, namely (a) it had to include a video game, (b) it had to include some kind of adaptation, and (c) this adaptation was based on emotion. The string for the first characteristic was based on common practice in similar studies (e.g. [40]) and included GAME* OR GAMING. The search string for the second characteristic was based on game adaptation literature that used synonyms for adaptation processes and included ADAPT* OR MODUL* OR ADJUST*. Lastly, the string based on the third characteristic was based on affective computing studies and terms used for emotions or emotional components, namely: AFFECT* OR EMOTION* OR VALENCE OR AROUSAL OR EXPERIENC*.

3.2.3 Inclusion Criteria

This review aims to investigate the reported effect of affective-adaptive games and how these effects are being empirically evaluated. Therefore, it focuses on high-quality comparative studies, leading to the following inclusion criteria:

1. Peer-reviewed papers (including conference papers)

2. Full-length papers
3. Available in English or German
4. Test an adaptive video game based on affective information
5. Evaluates the adaptation effects empirically against a control condition

3.2.4 Exclusion Criteria

Following the reasoning to provide insight into high-quality work in the field, studies were excluded if they showed one of the following characteristics:

1. Do not include an empirical study (i.e. reviews, study protocols, 'work-in-progress')
2. Evaluate only through qualitative or descriptive means
3. Do not compare to a control condition that is not affect-adaptive
4. Evaluate only based on case studies (defined as $N < 5$)

It is important to note that all non-affect-adaptive control conditions were included in the study, including performance-adaptive or non-adaptive games that were tested in a within-design. Additionally, it was not a necessary criterion to include randomized condition assignment, for example in quasi-experimental designs. Evaluations therefore did not need to consist of randomized controlled trials (RCTs) to be included.

3.2.5 Data Analysis

The initial search returned 3,930 papers, 755 of which have been identified as duplicates and were removed. Title and abstracts were screened by the principal investigator and papers that demonstrated a clear mismatch to any of the relevant research questions (e.g. papers that don't involve video games or HCI in general) were excluded, leading to the removal of another 2,965 papers. The 210 remaining papers were assessed by reading the full texts of the papers and coded in regards to the inclusion and exclusion criteria. Out of these papers, 32 were excluded for not involving an empirical study, 7 were excluded for not involving a video game, 38 were excluded for not involving an adaptation, 55 were excluded for not basing the adaptation on affective data, 36 were excluded for not evaluating the effects of affective-adaptation empirically, 14 were excluded for not involving a control

condition within the evaluation, and finally 2 were excluded for only evaluating through a case study. The final set of papers consisted of 26 studies that were further analyzed within this review. A full representation of the process as proposed by PRISMA guidelines [175] can be viewed in Fig 3.1.

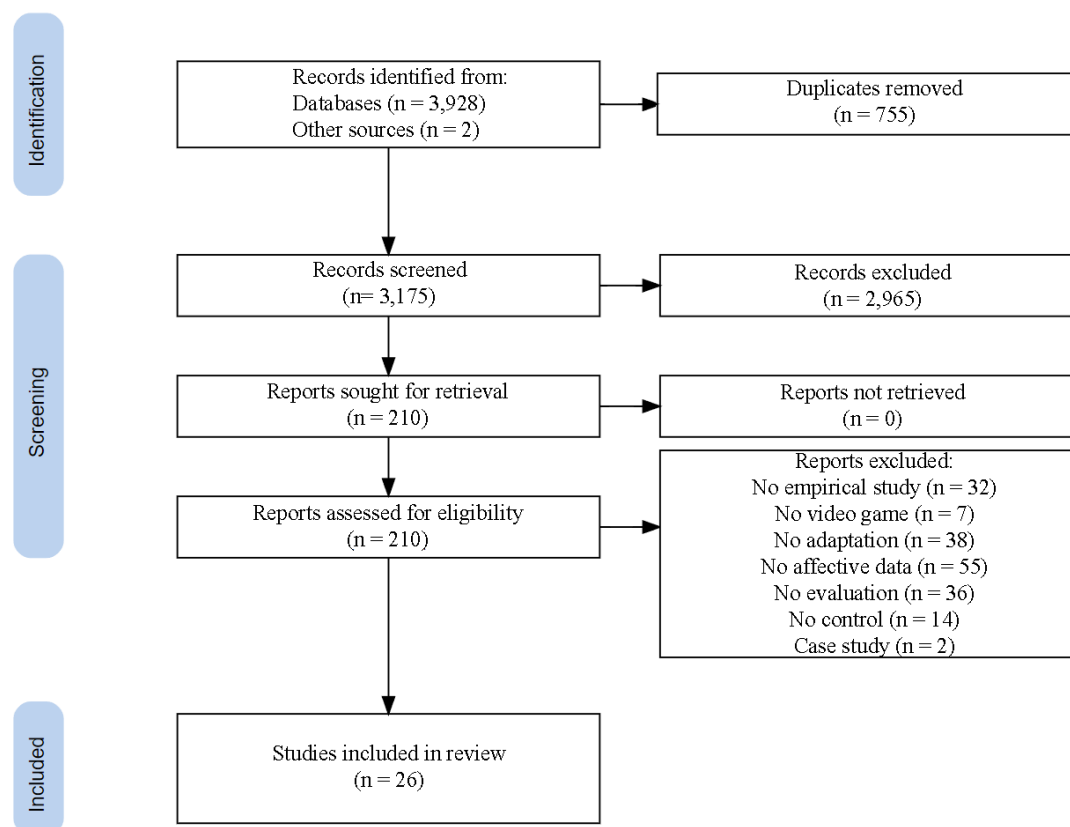


Figure 3.1: Flow diagram of data extraction process.

3.2.6 Coding

In order to answer our three research questions individually, specific aspects of the full sample were coded under predefined conditions for each question.

3.2.6.1 RQ1: Evidence Synthesis

Each study was coded by publication year and within each study, each adaptive game was coded by genre (as described in the paper itself). For each game that was tested, the specified outcome variable was coded, including the measurement instrument. The study design was coded based on the control being a within or between condition and the number

and nature of all tested conditions were specified. Finally, the effect was coded as positive, mixed/neutral, or negative (experimental condition compared to the control condition), and if possible the reported effect size was included and coded as small, medium, or high effects, based on interpretation guidelines as reported by [177]. For studies where no effect size was included, but sufficient data was provided, effect sizes were calculated and interpreted as described by [177].

3.2.6.2 RQ2: Theoretical Assumptions

For each paper, the affective state of interest (i.e. source of adaptation) was coded based on the theorized underlying structure (dimensional vs. distinct) and the reported labels of the measured emotional states. Furthermore, the specific measures used to detect the emotional state were recorded. Together these details were gathered in order to examine how affect was measured across studies. Each paper's efforts to validate individual measurement instruments (e.g. through comparison with self-report scales) were also recorded. Tests were either direct (i.e. related to subjective measures of the target emotion), indirect (related to other indications of target emotions), or absent.

Additionally, it was coded what game material was adapted to affective information and whether these game materials were tested in their ability to elicit a target emotion to inform the adaptation design. Game materials that were adapted for each game were listed and summarized where appropriate (e.g. "difficulty" for all individual gameplay changes that were made to increase challenge). Tests were again either directly (impact of materials was related to subjective measures of target emotion), indirectly (impact of materials was related to other indications of target emotions), or not conducted.

3.2.6.3 RQ3: Methodological Approach

Finally, for each evaluated game, the methodological approach was coded, including sample information (N, percentage of male participants, mean age) and the used statistical test. An estimate of achieved statistical power was calculated post-hoc for each study based on the study design, sample size, and an assumed medium effect size (0.5 standard deviations [SD]). Rather than providing an estimate of "achieved" power, this was done because such an estimate completely depends on the observed effect and can therefore be misleading, as it is not theory-based, nor a good indicator of methodological validity [178]. Additionally, many studies did not provide sufficient information to calculate the observed effect size,

which would limit the ability to compare all studies. To provide more insights about each statistical power, target effect sizes (ES) were calculated, representing the detectable effect sizes for a study, assuming a power (a priori) of at least 0.8. The target effect size therefore represents the necessary differences between groups in SD to achieve a power of 0.8 or higher.

Furthermore, risk of bias (RoB) was assessed using the Cochrane Collaboration's tool [179]. Risk of bias was assessed based on objective criteria regarding multiple domains: (a) selection bias (i.e. whether or not participants' allocation was concealed and randomized), (b) performance bias (i.e. whether participants were aware of the intervention and if this could affect outcomes), (c) attrition bias (i.e. how much missing data regarding the outcome was reported and how that could influence analyses); (d) detection bias (i.e. whether or not clear and appropriate measures for the outcome were reported and whether deviations arose through data collection strategies); and (e) reporting bias (i.e. whether or not all results from all measurements and analyses were reported). An overall RoB was judged based on the following criteria:

1. Low risk: The study presents a low risk of bias for all domains
2. Some concerns: The study presents some concerns in at least one domain, but no high risk for any domain
3. High risk: The study presents a high risk in at least one domain

A detailed overview of all domains and criteria was provided by [179].

3.3 Results

3.3.1 RQ1: Effectiveness of Affective Adaptation

26 studies were included in the analysis. A description of study aims, methods, and conclusions can be found in Appendix A.1. An overview of publication years can be seen in Fig 3.2. Over half of all studies ($n = 14$) were published in 2018 or later. 69% ($n = 18$) tested an affect-adaptive game against one or more control conditions in a repeated measure design and 31% ($n = 8$) used a group comparison. 86% of studies ($n = 22$) used randomized subject assignment, 8% ($n = 2$) did not use randomization for subject assignment, and 8% ($n = 2$) did not report sampling procedures.

In these 26 studies, 27 affect-adaptive games have been described. An overview of the games' genres can be seen in Table 3.1.

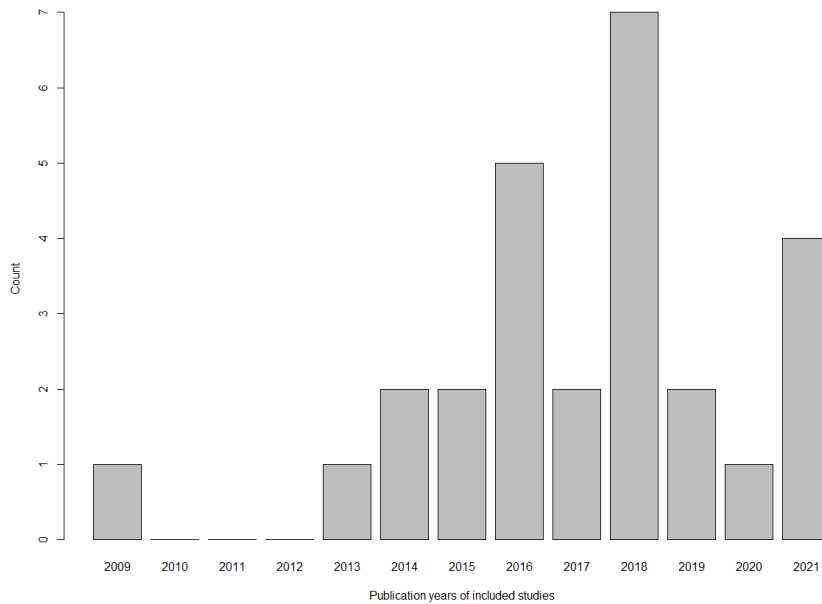


Figure 3.2: Counts of included studies by year.

To evaluate adaptation effectiveness, 18 different outcome variables were used, assessed through 15 different instruments (see Table 3.2). The outcome variables can be summarized within three broad categories. The most used outcome category ($n = 16$) is player experience, which includes outcome variables such as enjoyment, engagement, immersion, aesthetics, dynamics, competence, character believability, fun, flow, and general player experience. 46% of studies ($n = 12$) measured player experience through a previously validated self-assessment instrument, such as the Game Experience Questionnaire [180] (GEQ; $n = 3$), the Intrinsic Motivation Inventory [181] (IMI; $n = 3$), the Flow Experience Measure [182] (FEM; $n = 1$), the Player Experience of Need Satisfaction [16] (PENS; $n = 1$), the Immersive Experiences Questionnaire [183] (IEQ; $n = 1$), the Player Experience Inventory [184] (PXI; $n = 1$), User Response to Interactive Storytelling tool [185] (URTIS; $n = 1$), and the Character Believability Questionnaire [186] (CBQ; $n = 1$). Additionally, 27% of studies ($n = 7$) constructed their own scales to assess player experience.

Another category includes affective variables ($n = 8$), such as arousal, stress, valence, excitement, and anxiety. These were measured mostly through physiological data, including heart rate (HR; $n = 1$), heart rate variability (HRV; $n = 3$), electrodermal activity

Genre	No of studies	% of sample
Action (3D)	5	18
Arcade	3	11
Education	2	7
Horror	4	15
Interactive Story	1	4
Platformer (2D)	4	15
Shooter (3D)	4	15
Training	4	15

Table 3.1: List of genres for adapted games analyzed in this review.

(EDA; $n = 3$), and electroencephalography (EEG; $n = 1$). Some studies measured the affective outcome through facial expression recognition (FER; $n = 1$) or voice analysis ($n = 1$), and finally some through subjective self-assessment tools such as the Self-Assessment Manikin (SAM; $n = 1$), the Mood Adjective Checklist (UMACL; $n = 1$), or an own scale ($n = 2$).

The third category consists of performance metrics ($n = 9$), either in-game performances ($n = 8$) or learning performance metrics ($n = 1$).

Table 3.2: List of included studies, outcome variables, outcome assessment instrument, effect direction (non-significant effects marked with n.s.), and observed effect size if sufficient information was provided.

Authors	Outcome	Instrument	Control	Effect	Effect Size
[187]	Player Experience	GEQ [180]	Non-adaptive	Positive	N/A
[152]	Stress Reduction	Physiology (HRV)	Non-adaptive	Positive	Medium
[188]	Flow, Performance	GEQ [189]	Performance adaptation	Negative	Small
[190]	Valence Decrease, Arousal Increase	Facial Expression	Non-adaptive	Positive	Medium
[191]	Preference	Single Item	Non-adaptive	Positive	N/A
[192]	Effectiveness, Efficiency, Difficulty	In-game Performance	Non-adaptive	Positive	N/A
[193]	Effectiveness, Efficiency, Difficulty	In-game Performance	Non-adaptive	Positive	Medium
[194]	Player Experience	IMI [181], FEM [182]	Manual, Random, Performance, Personality adaptation	No effect	N/A
[195]	Enjoyment, Immersion	UMACL [196], IEQ [183]	Manual adaptation	No effect	N/A

Continued on next page

Authors	Outcome	Instrument	Control	Effect	Effect Size
[197]	Perceived Competence, Aesthetics, Dynamics	PXI [184], IMI [198]	Increasing difficulty, Fixed difficulty	Positive	Large
[199]	User Experience	URTIS [185]	Non-adaptive	No effect	N/A
[200]	Presence	SUS [201]	Non-adaptive	Positive	N/A
[202]	NPC Rapport	3-item questionnaire	Non-adaptive	No effect	N/A
[203]	Affective state, performance	Voice analysis, In-game-performance	Non-adaptive	Positive, no effect	Medium
[204]	Player Experience, Performance, Anxiety	Single Items (9-point Likert)	Performance adaptation	Positive	N/A
[205]	Challenge and Experience	IEQ [183]	Non-adaptive	Positive	N/A
[206]	Arousal, Player Experience	EDA, IMI [181] & PENS [16]	Non-adaptive	Positive, no effect	Large
[207]	Player Experience	GEQ [180]	Non-adaptive	Positive	Large
[208]	Physiological Arousal, Performance	Physiology (HRV, EDA), In-game-Data	Non-adaptive, Deep breathing task	Positive, No effect	N/A
[209]	Physiological Arousal, Performance	Physiology (HRV, EDA, BR), In-game-Data	Non-Adaptive, Deep breathing task	No effect	Large
[210]	Valence, Arousal, Dominance	Physiology (HR, EDA), Subjective (SAM [65], own scale)	Non-adaptive	Positive (n.s.)	Large
[211]	Performance, Flow	In-game Data, Single Items	Performance adaptation, Non-adaptive	No effect	N/A
[212]	Learning, Engagement	Not specified	Non-adaptive	Positive	Small, large
[213]	Long term excitement, Enjoyment	Physiology (EEG), Single Item	Non-adaptive	Positive	Small
[214]	Character Believability	CBQ [186]	Non-adaptive	Positive	N/A
[215]	Fear, Fun, Difficulty	5-point scale	Non-adaptive	Negative (n.s.)	Medium

While a variety of outcome variables were used, most studies reported a positive effect direction (i.e. increase in affect-adaptive condition compared to control). 65% of studies (n = 17) reported positive effects, of which 2 can be considered small, 4 can be considered

medium, and 6 can be considered large. 5 of these effects were not reported with sufficient data to calculate effect sizes. Only 4% of studies ($n = 1$) reported a significant negative effect. 42% of studies ($n = 11$) reported non-significant effects for at least some of their outcome variables.

3.3.2 RQ2: Emotions in affect-adaptive games

All presented games aimed at improving a predefined outcome variable by adapting game material to emotional states. They included means to measure affective states and emotion-eliciting material that was the aim of some form of adaptation. An overview of emotion-theoretical assumptions, emotion measures, eliciting materials and whether measures and material were tested in their ability to reflect target emotional states can be seen in Table 3.3.

Table 3.3: List of included studies, definition of underlying emotion structure, emotional state labels, emotion measure, whether the measure was tested in the study, the in-game adapted material used for emotion elicitation, and whether the effect of this material on emotion elicitation was tested in the study.

Authors	Structure	State Labels	Measure	Measure validated	Adapted material	Material tested
[187]	Distinct	Anger, Frustration, Smile, Relaxation	Facial Expression	Not tested	Difficulty	Not tested
[152]	Dimensional	Stress	Physiology (HRV)	Not tested	Visual Feedback	Indirectly
[188]	Distinct	Anxiety, Boredom, Engagement, Frustration	Physiology (HR, EEG)	Indirectly	Difficulty	Directly
[190]	Dimensional	Valence, Arousal	Facial Expression	Not tested	Difficulty	Not tested
[191]	Distinct	Anger, Disgust, Fear, Happiness, Neutral, Sadness, Surprise	Facial Expression	Indirectly	Difficulty	Directly
[192]	Distinct and Dimensional	[Anger, Disgust, Fear, Happiness, Sadness, Surprise] and [Arousal]	EDA and Facial Expression	Not tested	Difficulty, Lighting	Not tested

Continued on next page

Authors	Structure	State Labels	Measure	Measure validated	Adapted material	Material tested
[193]	Distinct and Dimensional	[Anger, Disgust, Fear, Happiness, Sadness, Surprise] and [Arousal]	EDA and Facial Expression	Not tested	Difficulty	Not tested
[194]	Dimensional	Preference	Physiology (EDA, EEG, HR, HRV)	Directly	Difficulty	Directly
[195]	Distinct	Boredom, Engagement	EEG	Directly	Cognitive Demand	Directly
[197]	Dimensional	Boredom, Frustration	Questionnaires	Directly	Difficulty	Indirectly
[199]	Distinct	Distress, Fear, Hope, Joy	Behaviour	Not tested	Narrative Trajectory	Not tested
[200]	Distinct	Anger, Disgust, Fear, Happiness, Sadness, Surprise	Gestures	Indirectly	Music	Not tested
[202]	Distinct	Alarmed, Angry, Bored, Content, Depressed, Happy, Miserable, Neutral, Tired	Physiology (EDA, EMG)	Not tested	NPC Dialogue	Not tested
[203]	Dimensional	Valence, Arousal	Voice Analysis	Directly	Difficulty, Sound	Partly directly
[204]	Dimensional	Anxiety	Physiology (HR, EMG, EDA)	Directly	Difficulty	Not tested
[205]	Distinct	Anger, Frustration, Joy	Facial Expression	Not tested	Difficulty	Not tested
[206]	Dimensional	Excitement	Physiology (EDA)	Not tested	Difficulty	Not tested
[207]	Dimensional	Arousal, Valence	Physiology (EDA, EMG, HR, HRV)	Not tested	Character representation, Difficulty	Indirectly
[208]	Dimensional	Arousal	Physiology (BR)	Not tested	Difficulty	Not tested
[209]	Dimensional	Arousal	Physiology (EDA, HRV, BR)	Not tested	Difficulty	Not tested
[210]	Dimensional	Arousal, Dominance, Valence	Physiology (HR, EDA)	Directly	Difficulty	Directly
[211]	Dimensional	Boredom, Frustration	Physiology (EDA)	Not tested	Difficulty	Directly
[212]	Distinct	Boredom, Frustration, Relaxation	SAM [65]	Not tested	Difficulty, Aesthetics	Not tested
[213]	Dimensional	Excitement, Frustration	Physiology (EEG)	Not tested	Difficulty	Indirectly

Continued on next page

Authors	Structure	State Labels	Measure	Measure validated	Adapted material	Material tested
[214]	Distinct	Anger, Disgust, Fear, Joy, Neutral, Sadness	Facial Expression	Not tested	NPC behaviour	Not tested
[215]	Distinct	Anxiety, Fear, Neutral, Suspense	Physiology (HR)	Directly	Enemy Position	Not tested

42% of studies ($n = 11$) considered emotions as distinct states, while 50% ($n = 13$) considered emotions as instances along a dimension. The remaining 8% ($n = 2$) explicitly defined and measured both distinct and dimensional affective variables. Adaptations were based on a wide variety of affective triggers that were often based on the means of measurement (e.g. arousal cut-offs with dimensional measures, and classified fear with distinct measures). The number of states that were measured ranges between 1 and 11.

Affective measures were used to indicate states by specific emotion component expressions. The most widely used form of measurement were physiological measures, used by 62% of studies ($n = 16$) and were conducted through HR readings ($n = 6$), HRV readings ($n = 4$), EDA ($n = 10$), EEG ($n = 4$), electromyography (EMG; $n = 3$), or breathing rate ($n = 2$). 42% of studies ($n = 11$) considered observational data of behaviours, such as facial expressions ($n = 8$), voice analysis ($n = 1$), gesture analysis ($n = 1$), or in-game choices ($n = 1$). Finally, 8% of studies ($n = 2$) measured subjective feeling in-game as a means to adapt gameplay through self-report ratings. 62% of studies ($n = 16$) did not explicitly test how well the used measure indicated target emotional states, meaning that these studies relied on either previously tested or untested theoretical assumptions regarding how well a measure could differentiate between predefined states based on a predefined underlying structure. 12% ($n = 3$) indirectly tested the measure, by validating it through other means than subjective emotion self-assessment (e.g. comparing physiological measures, or testing reliability of differentiating between emotion-eliciting game materials). 26% of studies ($n = 7$) tested the measure within a certain game context directly, by associating it with the self-reported target emotion in an experimental context.

The emotion-eliciting game material that was the source of adaptation was mostly focused on challenge aspects of games. 77% of games ($n = 20$) manipulated game material to change a game's difficulty in order to evoke a range of emotions. 19% of studies ($n = 5$) manipulated the game's aesthetics (through visuals or audio) as a way to evoke emotions.

15% of studies ($n = 4$) manipulated non-playable characters (NPCs) or story progressions to reflect affective data of players, and 4% of studies ($n = 1$) manipulated non-challenge related in-game events to evoke fear. Again, a majority of studies (58%; $n = 15$) did not test the effects of game material manipulation on the target emotion through self-report measures. 16% of studies ($n = 4$) tested the emotional elicitation effect of material through indirect measures (such as physiology), and 26% of studies ($n = 7$) tested the effects of the adapted game material on target emotions directly through self-report measures.

3.3.3 RQ3: Methodologies

The sample size (n) ranged from 9 to 294 ($M = 37.62$), with a Median sample size across studies of 24 participants. None of the studies justified the sample size on statistical power assumptions. 96% of studies ($n = 25$) provided information about demographic details, such as mean age ($n = 18$), age range ($n = 21$), gender distribution ($n = 24$), or game experience ($n = 15$). Statistical power assuming a medium (0.5 SD) effect size ranged from 0.08 to 0.99 ($M = 0.55$; $Md = 0.46$). The target effect size detectable with the study design ranged from 0.2 SD to 2.8 SD ($M = 0.89$; $Md = 0.8$). 8% of studies ($n = 2$) were able to detect a small effect size (up to 0.3 SD), 42% of studies ($n = 11$) were able to detect a medium effect size (up to 0.6 SD), 58% of studies ($n = 15$) were able to detect a large effect size (up to 0.9 SD), and 88% of studies ($n = 23$) were able to detect a very large effect size (up to 1.5 SD). An effect size of up to 2 SD was detectable by 96% of studies ($n = 25$) and one study was underpowered for lower effect sizes than 2.8 SD.

Table 3.4: Included studies, sample size (N), reported demographics (% male and mean age), statistical test, estimated power assuming a medium effect (0.5 SD), target effect size (ES), and risk of bias (RoB; + refers to low risk; +/- refers to some concerns; - refers to high risk of bias).

Authors	N	% Male	M Age	Test	Power	Target ES	RoB
[187]	60	68	N/A	t-Test	0.85	0.5	+
[152]	12	58	33.92	MANOVA	0.20	1.3	+/-
[188]	21	76	22.43	t-test	0.39	0.71	-
[190]	31	N/A	N/A	Wilcoxon Rank	0.84	0.5	-
[191]	25	80	N/A	Z-test	0.43	1.0	+/-
[192]	30	60	N/A	t-test	0.85	0.5	+
[193]	30	60	31.87	t-test	0.85	0.5	+
[194]	50	74	25.1	t-test	0.23	1.1	+
[195]	10	40	N/A	t-test	0.29	0.8	+

Continued on next page

Authors	N	% Male	M Age	Test	Power	Target ES	RoB
[197]	66	73	30	ANOVA	0.99	0.3	+
[199]	294	50	19	MANOVA	0.99	0.2	+
[200]	22	67	29.09	t-test	0.3	1.2	+
[202]	16	63	N/A	t-test	0.24	1.4	+
[203]	40	N/A	N/A	t-test	0.93	0.4	+
[204]	9	47	N/A	ANOVA	0.26	1.1	+/-
[205]	32	N/A	N/A	Wilcoxon Rank	0.85	0.5	+
[206]	16	94	N/A	ANOVA	0.61	0.7	+
[207]	24	67	22.5	MANOVA	0.41	0.8	+
[208]	9	78	N/A	Not specified	0.08	2.8	+/-
[209]	16	94	N/A	ANOVA	0.12	1.6	+
[210]	11	73	30.5	Not specified	0.19	1.7	+/-
[211]	36	61	N/A	Friedman test	0.82	0.5	-
[212]	30	67	19	t-test	0.38	1.0	+/-
[213]	24	92	25.59	ANOVA	0.74	1.0	+
[214]	52	86	N/A	Z-Test	0.97	0.4	+/-
[215]	12	92	25.42	Not specified	0.49	0.8	+

62% of studies ($n = 16$) were found to have a low risk of bias (RoB), i.e. no bias concerns in the observed domains. 27% of studies ($n = 7$) showed some concerns for risk of bias, and 11% of studies ($n = 3$) showed domains with a high risk of bias. Al Osman et al. [152] compared a biofeedback game against the same game with hidden feedback. They also introduced participants to the game aim and relaxation strategies through meditation in the biofeedback condition only. These conditions were therefore visible to participants and could impact the outcome, leading to some concerns in the domain of performance bias, even though the sampling was reportedly counterbalanced. Alves et al. [188] reported inconsistent empirical results (i.e. different effect sizes for the same effect), which was judged a high risk for reporting bias. Andrew and Chowanda [190] used strategies of un-concealed randomization, did not report group comparisons for all outcome measures, and proposed some conflicting operationalizations of similar measures (such as negative valence through FER and positive affect through self-report as desired outcomes), leading to a high risk in the domains of detection and reporting bias and some concerns in selection bias. Blom et al. [191] reported multiple outcome variables (preference, challenge, immersion, frustration) through constructed self-assessment questions, but only reported descriptive differences for challenge, immersion, and frustration without a statistical test to test these differences, which indicates some risk in reporting bias. Liu et al. [204] provided a clear methodology, but conducted some additional analyses and created variables not previously

justified, indicating some potential risk for reporting bias. Parnandi et al. [208] missed some important information in the process description (such as randomization, blinding, or how knowledge of different interventions [such as affective game vs. deep breathing exercise] was controlled in its potential to affect outcome). It is not clear if all outcomes are sufficiently reported, as a statistical test for group comparisons was not provided for all outcome variables, indicating some concerns about reporting bias. Rodriguez et al. [210] were not able to randomize participants across conditions, as experimental data was compared to a previously conducted experiment. Additionally, they provide very limited reports of group differences for all outcome variables, indicating some concern for selection bias and reporting bias. Rosa et al. [211] did not provide a clear analysis plan (including the number and types of outcome variables and statistical tests), leading to some potential replication issues and a high risk for reporting bias. Potential order effects due to missing counterbalance were not discussed, indicating some concerns for selection bias. Salah et al. [212] reported extremely large effects (> 5 SD group difference), without sufficient indications of the potential nature of these effects. Measures such as "learning effect" were also not clearly defined, indicating some concerns regarding performance and detection bias. Finally, Tjokrosetio and Chowanda [214] described an unconcealed randomization process, leading to some concerns about selection bias. Additionally, the outcome variable was tested by participants watching specific gameplay videos without playing the tested games, leading to an unclear evaluation of adaptation, as it was not described how emotion adaptation contributed to changes in outcome variables.

3.4 General Discussion

This study aimed to investigate the impact of affect-adaptive games on various possible outcome variables through a systematic review of high-quality evaluation studies of the field. To broaden our understanding of the nature of these studies, both theoretical assumptions regarding emotion research, and methodological concerns were examined. 26 studies were identified that evaluate emotion-adaptive games against a form of control condition in an empirical context and their contents were summarized.

3.4.1 The Effects of Adaptation

To judge whether or not emotional game adaptation can be seen as effective, there are many variables that need consideration. In the initial search, many studies were identified that describe methods to achieve affect-adaptive adaptation, but many did not focus on evaluation ($n = 36$), some did evaluate but either without a control condition or only using case studies ($n = 16$). Still, the empirical evaluation of affective games against controls has been a topic with increasing interest, as 26 studies were identified, most of them published after 2017. These studies test a range of different adaptation mechanisms in different genres of games, with different strategies to measure and model emotions, and even different outcomes of interest.

The most investigated outcome related to at least some domains of player experience (PX). As a concept, player experience suffers from the lack of a clear conceptualization and measuring standard, which was mirrored by the abundance of different instruments to measure PX aspects. Only recently efforts have been made to test and improve validity and reliability concerns. For example, Denisova et al. [216] tested the underlying structure of the IEQ, GEQ, and PENS and found considerable similarities, which makes a clear distinction between tested PX domains difficult. Similarly, Johnson et al. [217] tested the factor structure of the GEQ and PENS and found they were only partially replicable. Aspects of concepts like immersion and flow show considerable overlaps, leading to further doubts about how many and which domains PX consists of [218], influencing the value of PX as a precise and valid research outcome and therefore as a useful development concept. Because integrative and comparable research becomes more and more important to evaluate effects, my findings support the notion of the need of more unified concepts and instruments, especially in terms of game evaluations.

Still, using PX as a broad overall category of interest, mostly positive effects of emotion-adaptive games have been reported. For example, Akbar et al. [187] provided empirical evidence for PX improvements through DDA using facial expression recognition for both a 2D platformer and 3D shooter and similar results were reported by Moniaga et al. [205] for a 3D Hack and Slash game. Frommel et al. [197] used in-game self-reported emotions to adapt difficulty in a 2D platformer, leading to large effects. Nogueira et al. [207] extensively tested multiple versions of affective adaptation through physiological data in a survival horror game and identified many large PX domain improvements compared to a non-adaptive game. Ibanez et al. [200] showed improved presence for a virtual reality

horror game with fear-adaptive music against the same game with generic music.

There were some non-significant effects reported, which could indicate mixed results regarding the effect of adaptation. Many of these however also indicate small sample sizes and a low statistical power, making it difficult to draw inferences. Darzi et al. [194] for example found no effect on multiple PX domains, but included many conditions, which led to a power of under 0.8 for any effect smaller than 1.1 SD. A similar picture can be seen in the study from Ewing et al. [195], who did not find an effect against manual adaptation of difficulty, or Negini et al. [206], who found no effect for PX reports, both showed a generally low power. Jalbert et al. [202] tested rapport with emotion-adaptive NPCs, but also was severely underpowered for any effect smaller than 1.4 SD. While this does not necessarily mean that negative or non-effects are always based on power, it is very difficult to interpret results that are not sufficiently powered to uncover a range of effect sizes. The study by Hernandez et al. [199] provides an exception; they had a large sample size, but still found no effect of emotion-adaptation on PX. In this particular study, emotion was measured through the choice of in-game dialogue and classified based on designer-constructed rules, which introduces a range of validity concerns regarding whether or not the emotion-adaptive game could truly be considered emotion-adaptive (as this was not tested using any validated emotion measure).

One important aspect to note is that while issues in statistical power become immediately apparent in studies with non-conclusive results, there are also issues in studies reporting significant results. Because the observation of significant results with a small sample size means that the observed effect is quite large, a high post-hoc power can be misleading and should not be interpreted as strength of evidence [178]. In fact, most studies in this review only achieve a sufficient power with large (0.8 SD) or very large (> 1 SD) effects. Even if these are found, issues in generalizability due to the small sample size should be considered. Salah et al. [212] conducted a study with a low sample size and found an extremely large adaptation effect for engagement (> 5 SD). While it can be argued that there is no need for large samples if the theorized effects are large enough to be observable, a small sample is also less likely to represent a given population [178]. Extremely large effects for small samples might lead to unreliable interpretations as the same effect might not hold true for a general population. Studies of Al Osman et al. [152], Blom et al. [191], and Liu et al. [204] have similar issues and report large positive effects in at least some of the observed outcome variables with a low sample size. Statistical

power was not explicitly discussed as a factor to justify sample size in any of the examined studies, and neither was accuracy. It is important to note that accuracy (i.e. width of confidence interval) can be seen as a considerable concern with most of the studies (given the median sample size of 24), making even significant effects potentially unrepresentative [219]. Additional concerns regarding generalizability and replicability were the inconsistent reporting of basic demographic data and descriptive statistics.

Studies that focused on affective outcomes reported positive to mixed results. Lara-Alvarez et al. [203] provided evidence for successful improvements in experiences of pleasant-high affective states in an emotion-adaptive learning game using pre-validated voice analysis. Stein et al. [213] used an EEG-adaptive version of a 3D shooter and showed higher long-term excitement values compared to the control version of the game. Parnadi et al. [208], [209] showed mostly no differences, comparing a relaxation training game to a non-adaptive game and a deep breathing task condition with very low sample size, leading to a general conclusion that affective games have promise in their ability to manipulate emotions through context (e.g. the ability to create stressful situations), which cannot be done with regular relaxation exercises, but the proper design and development of affective games need further work to provide consistent results. Rodriguez-Guerrero et al. [210] tested an affective against a non-affective neurohabilitation game with a low sample size and found inconclusive results, indicating complex affective relationships between game materials, player data, and outcomes. Vachiratamporn et al. [215] tested the effects of a fear-adaptive horror game in terms of emotional reactions, which remained non-significant, possibly based on a very low statistical power.

Studies that focused on the effects on performance [192], [193] reported positive effects for shooting, puzzle, and exploration tasks in a 3D game for an adaptive game compared to a non-adaptive game, using physiological and face recognition information. In these particular studies, it is argued that the combination of relevant information (in this case affective information and playing style classification) to personalize experiences could lead to the largest effect. The authors conclude that there are still many unknown variables and interactions when it comes to affective adaptation, but the initial promising data points towards the potential of further research, especially research that reduces the cost and obtrusiveness of affective recognition and modelling.

All in all, there is a lot of variance in multiple aspects of the analyzed studies, which makes a clear picture of the effect of emotion adaptation not yet possible. In fact, the

differences in methodological approaches, efforts to ensure generalizability, and efforts to reduce risk of bias add to the already present problems of comparability. Meta-analytical strategies, which are seen as one of the best ways to aggregate scientific knowledge [220] are difficult to conduct, not only because of differences in approaches and theoretical perspectives (such as outcome variables of interest or emotion models), but also because of differences in methodologies that should be universally prevalent, such as shared and precise PX conceptualizations, appropriate measures, well-constructed and powered experiments, and the sufficient reporting of data. While the reported effects of affective games seem ultimately promising, it may be too soon to fully evaluate them, given these barriers.

3.4.2 The role of Emotion

All studies described games that adapt material to affective information, which was either continuously or intermediately measured. In general, affective states of interest can be considered emotional, i.e. states with a relatively short duration and high intensity. While the elicitation of certain moods (e.g. in horror games) was a particular aim, all studies measured and adapted to data relating to emotional reactions, either measuring physiological aspects (through HR, HRV, EDA, EMG, or EEG), behavioural aspects (through FER, gesture analysis, or voice analysis), or subjective feelings (through self-reports).

Mostly depending on the measurement instruments, the inferred emotional states are either considered as dimensional or distinct constructs with states of interest that are considered useful for a particular game adaptation. For example, some studies [208], [209] focus on emotional arousal, measured through physiological arousal in an effort to create games for relaxation training. Others [187], [191]–[193], [214] use facial expression analysis to measure distinct emotional states, such as fear, joy, anger, or sadness. One of the main concerns when it comes to emotion measures is the inability of a single instrument to accurately reflect the complex nature of emotion in its entirety [38]. Inferences made from one or multiple measures are also subject to different sources of variation, such as personal differences and current context [32], [221]. That means that the validity of the emotion recognition system is highly dependent on the following factors: The measurement instrument, the emotion conceptualization, the given context, and individual differences. As the study by Rodriguez-Guerrero et al. [210] shows, even well-established emotion concepts (in this case dimensions of valence, arousal, and dominance), measured through a combination of instruments (such as HR and EDA), can lead to poor accuracy. Still, the majority of

studies ($n = 16$) did not explicitly test how well a certain measure predicted the target emotional state and, instead, built the emotion recognition system on theoretical assumptions. While some of the assumptions have considerable representation in the literature (such as the association between physiological and emotional arousal [145]), others are highly contested. For example, there is no clear consensus on which true emotional states are represented well through facial expressions [222]. Researchers (e.g. [187], [205]) may therefore interpret potentially non-distinct facial expressions (such as smile and smirk) as distinct emotional states. Another contested point is how and if distinct emotional states could be mapped to affective dimensions (e.g. [190]), as dimensional and distinct theoretical frameworks of emotions often have vastly different theoretical bases [28]. Finally, the exact relationship between a physiological measure and an affective state is not clear for every individual and context [32], so the relationship is hard to interpret without concrete mappings that some of the studies did not provide [192], [193], [202], [213]. While basing decisions on contested assumptions can be in some cases useful, especially in providing more insight into fundamental psychology research, without explicit validity testing, there is a risk in unknowingly misinterpreting ambiguous data. In the study by Ibanez et al. [200], the gesture-based emotion recognition was tested indirectly by classifying participants who encountered a predefined “emotion-inducing” room within the game world and accuracy was only sufficient to distinguish between participants who visited the fear room or participants who visited any other room. Given a specific game and audience, such an approach could provide a way to adapt between two affective states, although it is unclear if these states truly represent the targeted fear vs. no fear states. Alves et al. [188] combined measures indicating fear and frustration into a combined emotional state to increase accuracy, although the theoretical and practical implications of such a state are not discussed.

The explicit (and direct) testing of measures given a game context and player base has in some of the analyzed studies been used to improve emotion recognition strategies: Ewing et al. [195] described a 2-step process, first establishing a relationship between measures and target emotions and then designing adaptations. Frommel et al. [197] measured the feeling of a target emotion through self-reports, which directly reflected the base of potential adaptation. Liu et al. [204] based their study specifically on anxiety and established methods to accurately predict anxiety in a preceding experiment. To ensure theoretically valid mappings, the relationship between a proposed emotion model and a

given measured emotion component not only ensures valid predictions but also provides the opportunity to focus on any emotional state that might be of interest for game design, including complex emotions like shame or pride. In this sense, designers are not limited to measuring concepts with more established physiological correlates (such as emotional arousal), especially given the influences of context and individual differences that justify testing in any case.

Still, emotion recognition is only a part of the adaptation process. A game is only truly adaptive if it changes in a way that elicits a target emotion, which closes the feedback loop [27]. Again, most investigated studies make theoretical assumptions regarding such an elicitation process. Most notably, many studies propose affective difficulty adjustment based on the flow model [133], [134], which proposes the existence of an optimal experience (lying between dimensions of boredom and frustration) when challenge and skill of a game are balanced. As a consequence, many of the studies chose to adapt difficulty aspects of games (such as health, enemy behaviour, platform size, game speed, etc.) to achieve an optimal experience. But not only is flow a conceptually ambiguous construct in psychology [223], the precise relationship between skill, challenge, and flow is unknown [224]. Furthermore, it is hard to assess whether or not a given adaptation was successful if fundamental and untested assumptions must be made (e.g. smaller platforms lead to challenge, which leads to frustration). Again, as with all emotional reactions, elicitation has been found to be dependent on individual and contextual factors, both in perspectives that argue for basic, innate emotions [34], and in perspectives that argue for constructed emotions [38]. This is not only true for ambiguous concepts such as flow, but all emotions. Hernandez et al. [199] based their adaptation purely on the designer's ability to infer emotional states from made choices, which might have led to the observed lack of adaptation effects. Ibanez et al. [200] assumed specific relationships between game elements and six basic emotions based on Ekman [56] (e.g. light and flowers for joy, insects and slime for disgust) and used these assumptions to train emotion classifiers.

Again, the explicit (and direct) testing of eliciting material given a game context and player may be necessary to avoid unclear mappings between game materials and emotions. For example, Darzi et al. [194] tested the ability of different game characteristics to elicit the targeted emotional changes in a preceding test. Such a process could provide similar benefits as testing the relationship between emotions and measurement instruments. Moreover, if the relationship between the target emotional state and the game material is clear,

adaptation can be based on very specific, pre-defined rules that are not based on potentially contentious assumptions and address concerns of interindividual and intraindividual differences in emotion processing.

Overall, the nature of emotions given modern theories, including emotion component expressions, and the implications of theoretical perspectives are not thoroughly addressed in almost all studies, leading to potential theoretical uncertainties, and influencing the observed results.

3.5 Conclusion

This review provided aggregated evidence regarding the effects, evaluation methods, and theoretical assumptions of emotion-adaptive video games. Not only were mixed effects observed in the investigated studies, but a large variance in methodological approach and theoretical justifications was found, leading to many open questions regarding affective games. This systematic review adds to the body of evidence uncovering gaps in research and practice when it comes to games that adapt to player emotions.

Many of the analysed studies describe their main contribution as the development and exploration of technological solutions regarding emotion recognition and adaptation and not as the evaluation of affect-adaptive games. This review specifically analyzed the evaluations in terms of emotion-theoretical assumptions, methodologies, and findings. From such a perspective, it is clear that more work is required to draw certain conclusions regarding the three main aspects of affective gaming as defined by Hudlicka [142], i.e. emotion sensing, modelling, and adaptation. The main limitation for conclusive data in the field may be rooted in technology, but rather in theoretical and methodological standards. The research criteria regarding adaptation evaluation shared between studies are limited, especially in regard to generalisability. Ambiguous constructs, measured through instruments with unknown reliability are often used as outcome variables to evaluate adaptive games and there is a large variance between studies that makes it apparent that these constructs (especially relating to player experience) need more unification. Similarly, emotion-theoretical details are insufficiently integrated into the research process, leading to potentially erroneous practices in regard to applying emotion theories. The strongest support for the potential of affect-adaptive games in enhancing player experiences, performance, or health lies in studies that specifically test their affective assumptions in terms of (a) measures of affective data and their relationship to the target emotion; and (b) adapted game materials used to

elicit emotions and their relationship to the target emotion. Following such a process gives game designers and researchers the opportunity to gather more information, address concerns regarding the influences of individual differences and context on emotional reactions, and avoid making assumptions based on contentious emotion-theoretical perspectives. As many of the described studies show, emotion adaptation is promising if the design and evaluation process is robust. The evidence does not suggest that this potential holds only true for certain game genres, certain emotions, or certain measurements, but for a variety of games with different aims and potentially a variety of emotions, especially under-explored and complex emotions, measured through different pathways such as inferred through subjective, behavioural, or physiological data. The field of affective gaming holds many yet unfulfilled promises not only to enhance games but to further our knowledge of what emotions are. But - with the core issues analyzed - the path to contributing potential solutions is uncovered.

Chapter 4

Narrowing the Gap

"It is science's methodology to try to reduce complex phenomena like emotions to a list of functional requirements, and it is the challenge of many in computer science to try to duplicate these in computers to different degrees, depending on the motivations of the research."

— Rosalind W. Picard [225]

Following the basic findings of Chapter 3, there are two particular gaps that need urgent attention:

First, there is a need for a theoretical synthesis of the relevant fields. While psychological research provides us with a great number of theoretical concepts and empirical evidence for these concepts, these findings are not often put to use in game research with all their implications in mind. Emotional game studies are often only indirectly related to modern emotion theories and affective game studies suffer from multiple measurement and interpretation issues that are partly dependent on their assumptions regarding underlying emotion theories [1], [44]. Because emotions are such a complex psychological construct, there is currently no perfect solution to integrate findings from psychology, design, and computer science into a work that would be able to perfectly measure and interpret affective data and further use this data as a base for video game adaptation. Rather, integrative theoretical work should be able to provide a robust model of emotional interactions between humans and video games in a way that is applicable to all kinds of games, to all kinds of research and design aims, and even to all kinds of contradicting emotion model assumptions, as there is as of yet no universally agreed upon emotion model and basic conceptual aspects

of emotions are still in need for further research. In light of this situation, it might not be the most rewarding approach to try to explain this complex relationship in all aspects but to focus on aspects that are (a) known and generally agreed upon by experts in the field, and (b) applicable and therefore useful in making practical decisions.

Second, given a theoretical model that fulfils the above-mentioned criteria, there is furthermore a need for more validated and data-driven practical application of theoretical assumptions regarding emotions in video games. As it has been shown, it is quite tempting to make certain assumptions regarding emotional game aspects and player behaviour, but leaving them untested can prove risky and further widen the gap between theory and practice in the field. In fact, precisely because the theoretical base of the player-game relationship is so complicated and incomplete, practical methodological contributions become much more important. Giving researchers and game designers the opportunity to focus their efforts on gaining new empirical insights would prove valuable in multiple aspects. For example, game designers should be able to make emotion-adaptive games, even without a thorough understanding of emotion theories. Researchers on the other hand should be able to study specific emotional phenomena using games, without relying on already established measures or models with perfect fits (as those still do not exist). Therefore, providing a practical guide to establish, research, and use emotional relationships between players and games would be a valuable contribution to the field.

To summarize, this chapter will focus on the development and discussion of a framework for affective interactivity, consisting of two parts:

1. **A theoretical overview.** The work will integrate relevant literature in emotion research, design research, and affective computing and use this to develop a theoretical model based on the ongoing emotional interaction in video games. This model will be robust against theoretical uncertainties and can be used by researchers and developers to describe and model the emotional relationship between players and games.
2. **A practical guide.** To bridge our theoretical knowledge with practical methods that allow the development and research of emotional games, a methodological framework will be presented, discussing steps that enable researchers and designers to empirically test emotional relationships within the player-game interaction and make informed decisions based on these relationships in order to adapt the game to affective information.

4.1 Improving the Theory

As illustrated in Chapter 2, a lot of progress has been made in framing the emotional relationship between players and video games. Here we describe a theoretical model with clear definitions for each component and therefore integrate contemporary research from psychology, game design, and affective computing into one interactional model. This model expands the version of the affective feedback loop as described by Yannakakis and Paiva [47], specifically by implementing an emotion-theoretical model of the player into the loop and providing new labels to more precisely talk about emotional information that is being exchanged within the game-player interaction.

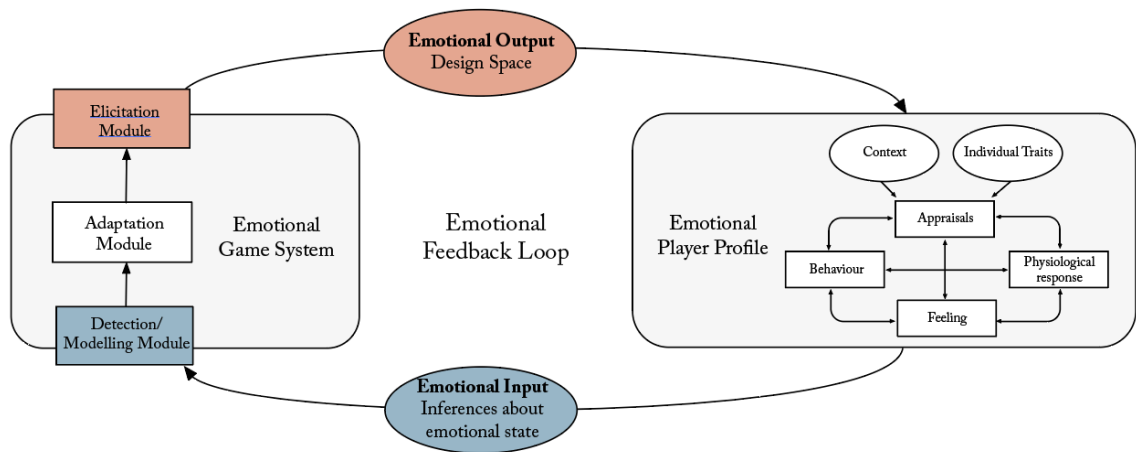


Figure 4.1: Illustration of the emotional feedback loop as an expression of game-player interaction. The game system uses emotional input that infers information about player emotions and adapts the emotional output accordingly. The player appraises the output on multiple variables which leads to interactions on measurable emotion components that can in turn be used to close the feedback loop.

4.1.1 The Player

The new player component in the context of emotional games specifically means the player's emotions. Depending on the design perspective, emotions are sometimes defined as experiences, sometimes as action motivations, and sometimes as bodily reactions. Drawing from shared assumptions in emotion research, we can assume that (a) emotions consist of multiple components (like physiological, behavioural, and subjective feelings); (b) appraisal

is central to the experience; (c) appraisal and the experience are dependent on context; (d) appraisal and the experience are dependent on individual differences; and (e) emotions evolve over time.

With this definition as a base, we are able to illustrate the player as part of the ongoing feedback loop with the game, adding specific emotions as psychological targets to the affective feedback loop that already exists in HCI studies [27], [42]. Emotions are here defined as short and intense affective states to reduce the inconsistent and imprecise use of the word "affect", which is especially important given the variety of affective feedback loops present in the literature. This model therefore represents an *emotional* feedback loop (**Fig 4.1**). To be precise, it explains relevant concepts of the affective relationship between player and games in the domain of emotions and not other affective states, such as moods or traits.

The advantage of viewing the player component as a direct representation of theoretical overlaps between emotion theories lies in the flexibility of the approach. Player emotions can be meaningfully interpreted as distinct categories, dimensional expressions, basic and universal mechanisms, as well as culture-dependent constructs. Either way, emotions are represented within multiple components of interest, such as physiological reactions or behaviours. One of the most important components in explaining emotion elicitation is the appraisal component, which is influenced by both individual traits and the situational context. Acknowledging the importance of traits and context within this model is an important step in not falsely generalizing the effects of certain stimuli on player emotions. Finally, because emotions consist of different qualities, it is also important to not view them as fixed, but dynamically evolving states. With consideration of all these aspects of human emotions, a useful representation of player emotions can be modelled and analyzed.

Still, because this player component is only based on shared theories between different theoretical perspectives, it is by no means an exhaustive model of all aspects of human emotions. Instead, it should be seen as a simplified representation of any given emotion of a player and considered more in its utility in applying these concepts to real games without making false assumptions than in its accuracy in explaining every relevant mechanism relating to emotions.

4.1.2 The Game

In the discipline of affective games, a game system has per definition three major tasks it has to fulfil to become affective: emotion sensing and recognition, computational models of emotion, and emotion expression or adaptation [41]. Based on these tasks, Yannakakis and Paiva [47] provided descriptions of three game system modules: an emotion detection module (a module to measure and model player emotions), an adaptation module (a module to adapt the game world to these player emotions), and an elicitation module (a module to elicit target player emotions). These modules represent the game side of the emotional feedback loop.

While these modules might not be explicitly described in traditional games, they can be used to describe familiar functions of any game system that relate to player emotions. For example, if a game offers a hard challenge, it has the potential to elicit feelings of satisfaction, accomplishment, and pride (emotion elicitation). There is however still the risk of evoking unwanted frustrating experiences, therefore most games gradually increase the challenge based on prior performance or player choice (emotion adaptation). While this adaptation is not based on primary emotional data, it could be based on for example difficulty settings, i.e. subjective preference (emotion detection). Every game that builds an emotional relationship can therefore be viewed within the emotional feedback loop.

4.1.3 Closing The Loop

Given the game system and the emotional player profile, there are now two kinds of information that are being exchanged. First, there are the stimuli provided by the game for the player. This is the design space of the game, which includes possible stimuli that elicit certain emotions. The term *emotional output* can be used to label these stimuli. This can include any type of game stimuli, such as NPCs [226], avatars [227], game world, mechanics and narrative [47], graphics and music [228], but also more high-level concepts like rewards, creativity, and player agency [131], [229].

This specific concept was chosen to describe game aspects that have a relationship to player emotions. This relationship cannot be seen as universal, which means it is dependent on individual differences and context. Following the loop, the player expresses the emotional reaction in various components (such as feeling, physiological response, or behaviour). This expression includes information provided by the player to be measured through the game to infer emotional states. We conceptualize this information as *emo-*

tional input. Naturally, emotional input can consist of direct emotion measures, like in biofeedback games. However, physiological measures cannot be assumed to map to certain emotions [32] and given the complex nature of emotions and the varying experiences games want to achieve, input measures need to be considered in terms of accuracy, cost, effort, and usefulness [145]. Therefore, other information might be more appropriate to use, like indirect measures, which could include any input a player makes in a game, from button press patterns to in-game behaviours [230]–[232]. Furthermore, building on our knowledge of how emotions occur, inferences could also be made from assumptions about contextual aspects, individual player traits, and emotion progressions that relate to the target emotion.

Again, this is conceptually true for every game, meaning that every game has the potential to gain access and make use of emotional information. However, the emotional reaction of a player is very subjective and dependent on context and individual traits, which makes the emotional relationship as a whole dependent on the type of game and the given player profile. In other words, there is no universal answer to the following questions: (a) How and to what extent can a given emotional output evoke a target emotion? (b) How and to what extent can we infer the emotions of a given player through a predefined emotional input? To make practical use of this model, these questions have to be answered for each game individually. If these questions are answered, games can be specifically designed to target affective reactions and also adapt to changes of player emotions.

To summarize, the emotional feedback loop reflects the ongoing emotional interaction between a game component based on affective game systems [47], and a player component based on current psychological emotion concepts that are relevant, resilient to changes in our understanding of emotions, and applicable across emotion theories. The emotional interaction is an explicit focus point and is represented through emotional output and emotional input as sources of affective information. It can be used to describe every game, but in order to apply it, it is necessary to make the emotional information measurable and usable in a practical process.

4.1.4 Summary

To summarize, the emotional feedback loop provides terms to describe the emotional game system (and its tasks), the emotional player profile with relevant emotion-theoretical components, and the information that is being exchanged between the game and the player.

This model can be seen as an extension of the affective feedback loop [27], [47] to provide (a) clear terms that are consistent with their definitions in their respective fields, and (b) a comprehensive illustration with integrated, contemporary knowledge regarding emotion research, game design, and affective computing. This model can be used to more easily talk about aspects of the emotional game-player interaction. It is therefore useful in terms of describing functional designs for existing games, guidelines for developing emotional games (based on the knowledge we currently have), and researching emotional phenomena that are part of such an interaction.

For example, using this model, we can summarize the tasks of designing emotional games as follows:

1. The **design of emotional game output**: The creation of game stimuli that elicit a certain target emotion through appraisal, given a certain context and individual differences
2. The **integration of emotional game input**: The measuring of certain emotional expressions from the player that can be used to adapt the game in favour of the target emotion

In order to contribute beyond research integration and new terminology, the descriptive tools provided by the model can now be used to propose a framework for designing and researching emotional games.

4.2 Improving the Practice

The aim of this section is to provide a practical guide for game designers and researchers to develop and test video games or video game characteristics that are adaptive to player emotions. It builds upon the emotional feedback loop to address some of the issues related to emotion elicitation in games and streamlines the process of integrating emotional adaptivity into game projects. Building on other interactional design frameworks [121], this process will be referred to as **emotion design for video games**. To provide the most useful overview of possible methods to apply the emotional feedback loop, it is important to further avoid confusion with other subtly different concepts and accurately describe the individual process steps.

The practical framework is built upon the new emotional feedback loop model and has the aim to enable designers and researchers to make and test emotional hypotheses

themselves and use the resulting data in a “bottom-up” design process. The framework’s steps were based on standard practice in usability testing [233], [234]; i.e. (a) identifying a primary goal for testing; (b) observing and recording users; (c) analyzing data and proposing changes. Each step was expanded upon and refined by adapting the process to the current emotion research advancements identified in the model in Chapter 4.1.

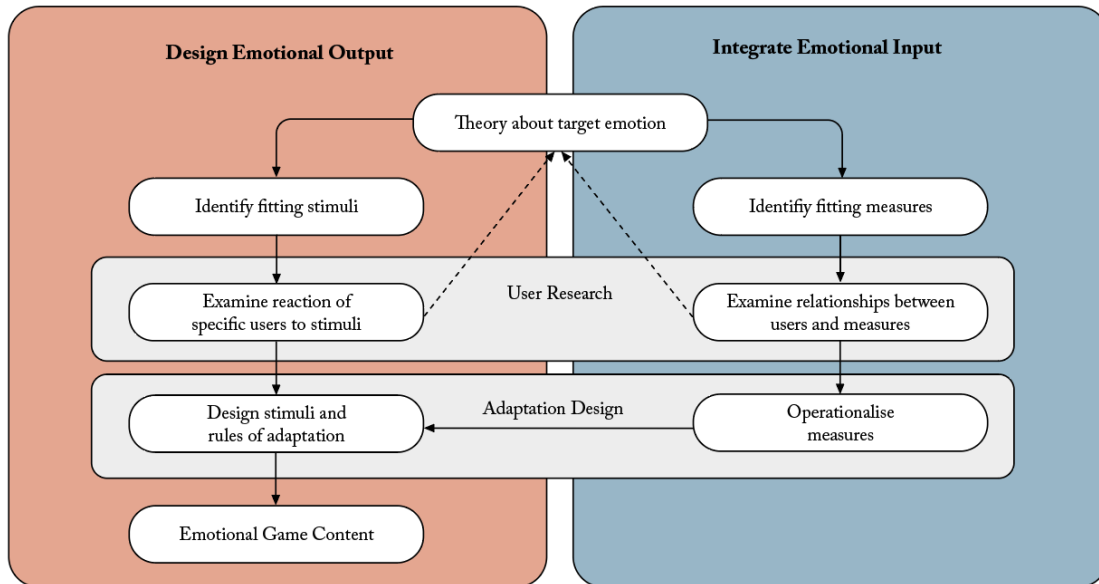


Figure 4.2: Illustration of the practical framework: To elicit a target emotion, both the domain of the game material and the domain of the user input need to be considered. Potential game characteristics (domain 1) and user input variables (domain 2) that could relate to the target experience need to be identified (step 1), and tested in user research (step 2). The gathered information can be used to introduce adaptivity of the game material based on user input to bring the product closer to the intended experience (step 3) or to improve research by further developing theoretical assumptions about the target emotion (dashed arrows)

4.2.1 Theory about Target Emotion

In order to design a game or game aspect that elicits a certain emotion, the target emotion in question needs to be conceptualized. In game design practice, the affective reward of games is often referred to as the experience [21], which can be considered as the subjective component of a certain emotional reaction. For example, the emotion "sadness" includes

as previously discussed many different components (e.g. cognitive, physiological, and behavioural nature), but the target itself for the designer is the subjective, conscious feeling of sadness. When choosing a target emotion, we can therefore at this stage consider only the subjective feeling a designer wants to elicit in a player.

Naturally, there is a growing interest in experiences that not only relate to broad "basic" emotions (e.g. fear or happiness) but also more complex and underexplored emotions [235], including "aesthetic" emotions (e.g. *schadenfreude* or the feeling of being moved [236]); or even emotions that have no concrete label (e.g. the anticipation of something bad happening or ordinary, less intense emotions [237]). Considering the labels of emotions also often differ between languages and cultures [31], the choice of target emotion should not be limited to well-known labels, but might also include experiences that are not yet fully understood or defined. Successful emotion elicitation might be facilitated if the target emotion has already been thoroughly explored in psychological research, but the only necessary characteristic of the target emotion is that it can be evaluated in terms of how successful the elicitation is.

When a target emotion is chosen, the subsequent process can be considered from two domains. The domain of emotional output tries to identify game material that evokes the target experience. The question that guides the steps in this domain is: What game characteristics evoke the target emotion? The domain of emotional input on the other hand tries to identify measures a game can use as an indication of the player's emotion. The question that guides the next steps in this domain is: What can a game track that indicates the target emotion? The framework introduces the terminology and process steps for these two domains for two reasons: (a) both domains need to be considered to ensure the target emotion can be elicited across individuals, and (b) both domains need to be considered to ensure the target emotion is elicited for one player during the progression of the game.

The framework specifically addresses uncertainties regarding emotion expression, differences in emotion elicitation depending on individuals and context, and the complex progression of emotions.

4.2.2 First domain: Design Emotional Output

4.2.2.1 Identify fitting stimuli

The literature describes many possible game elements that evoke certain emotional reactions, for example NPCs [238], avatars [227], game world, mechanics and narrative [47], graphics and music [228], but also more high-level concepts like rewards, creativity, and player agency [131], [229]. Frameworks like Emotioneering [130] or the mechanics–dynamics–emotions (MDE) framework [49] provide a variety of possible channels to use in the game design process.

It is however crucial to acknowledge that the characteristics of a stimulus are only important in the context of appraisal: It is not a fixed attribute of the dark hallway to evoke fear. Considering the classic appraisal variables described by Lazarus [89], goal-relevant but incongruent stimuli that are difficult to cope with correspond to fear, so the appraisal of the dark hallway as scary is dependent on specific features: does the player need to follow the hallway to reach a goal? Do players feel like the darkness is incongruent with their goal?

At this stage, game materials can be designed based on assumptions or existing examples indicating their potential in eliciting the target emotion. While it can be helpful to review known appraisal variables for well-researched emotions, the more practical and common approach is to design based on successful examples. For example, in an extensive analysis of the emotional effects of *Silent Hill* by Perron [150], the author discusses how the specific type of emotional terror in the *Silent Hill* series was designed to mimic Japanese psychological horror movies and how it differs from the experiences of other horror media. This discussion is led by appraisal on a reflective processing level. For example, during the production of *Silent Hill 2*, the developers tried to induce anxiety gradually increasing conscious anticipation of negative events [150]. Other processing levels can be considered, including sensorimotor responses (e.g. fear through sounds [239]), and perceptual and associative responses (e.g. fear through danger-associated stimuli).

4.2.2.2 Examine reaction of specific users to stimuli

Studies about industry practices show that game companies often rely on genre conventions, historical successes of other games, or internal theories of the developers when designing game stimuli [129], [235] and to great success. As observed through the lens of Uses and

Gratification theory, players too make use of genres, their knowledge of other games, and their own internal theories to predict their own feelings when playing a specific game [240].

But to make sure the game (or game elements) really evoke the target experience, game creators need to test them and this is commonly done through user research to thoroughly understand the individual emotional reactions of users within a target audience. In a 2018 discussion of game user research [241], Zammito of Electronic Arts states that almost all large game companies have teams of dedicated user researchers who are responsible for executing UX tests of a qualitative or quantitative nature to inform developers of the experiences of target audience members during various cycles of game development. Through organized processes involving design, development, and testing, many game aspects are iteratively fine-tuned to deliver on the target experience [242]. Within the domain of emotional output, user research can establish whether a game aspect can elicit the target emotion for certain individuals and in certain contexts.

The methods to conduct such research are often discussed in literature about player research or targeted empirical tests and usually include self-report instruments, such as for example questionnaires (e.g. the PANAS [60]), where players rate how strongly they felt an emotion in a certain moment of the game after playing. This provides a precise quantitative measure of experiences that is easy to interpret but may be influenced by restrictions of the questionnaire (e.g. not the complete emotional space is included). To capture a more complex view of the experiences, interviews or open questions can be constructed where players describe the experience in their own words (see for example [122]). This has the potential to provide a more detailed overview of the elicited emotions but might be harder to interpret for the use within a game. Additionally, self-reports after gameplay sessions could be influenced by factors like memory effects and only capture the most notable emotional reactions. To account for the evolving nature of emotions with variable onset and offset periods and subtle differences over time, continuous experience reports could be used (such as the with two-dimensional emotion-space [243], the affect rating dial [244], or even free association techniques [245]). Players get the opportunity to report their experiences continuously during gameplay and data can be collected that map exactly to in-game moments. Again, responses are limited to how open the scale is and free responses are more difficult to interpret and quantify. There is an exceedingly large space to describe experiences both in categories and finer gradients [246], so the choice of research methods should account for the needs of the specific designs.

4.2.2.3 Design stimuli and rules of adaptation

As mentioned before, games and specific game aspects are most commonly designed through iterations that are being evaluated in their potential to elicit a target experience. Using the proposed terminology of the framework, we can say that game stimuli are evaluated by their potential to become emotional output. This implicitly means for the designer that there is an at least partly known relationship between the game they create and the emotion a user experiences when playing it. Broadly speaking, this can mean that a game consists of similar elements that have been successful in other works, delivering an experience players can expect and seek out, but it can also mean on a detail level that specific aspects of a game have been extensively tuned and tested to ensure a certain emotional experience, even when played by a variety of players.

This by itself is not a new process, but we argue that given the complexity of emotion elicitation and expression, two criteria for the success of creating emotional game material should be considered at this stage: (a) how well does this design work across individuals; and (b) how well does this design work given the rapidly changing emotions within one individual.

We call the process that aims to ensure that these questions are addressed adaptation. Interindividual adaptation attempts to create the same emotional experience for different kinds of players. For example, players with a high ability to play a certain genre may enjoy more challenging gameplay, while players with lower ability would experience the same enjoyment with less challenge. Ideally, user research has identified game characteristics (such as reaction time windows or health for challenge) that are the source of different experiences for different types of players, so these characteristics can be accounted for (e.g. through purposefully designed difficulty settings).

Intraindividual adaptation attempts to create an optimal experience for one player as the game unfolds, based on assumptions of how emotions evolve. For example, a game has successfully elicited fear for a player, which means that the design space is now open for more intense fear reactions (e.g. terror), or strong opposite emotions (e.g. the feeling of relief).

If a design can elicit the target emotion across different individuals and within one individual, it can be considered adapted to player emotions and therefore as emotional output. We can illustrate this with the example of game difficulty: It is widely acknowledged that game difficulty has an influence on enjoyment when player abilities are taken

into account (e.g. in dynamic difficulty adjustments or in player-chosen difficulty settings) [247]. Many games provide difficulty options to ensure enjoyable experiences for different types of players but also adapt game difficulty during the game to accommodate the changing abilities of one player. In order for this to work, game designers need to be aware of the relationship between difficulty settings and enjoyment, i.e. emotional output and target emotion as discussed in the previous sections. But they also need to be aware of the current state of the player, which the framework addresses in the second domain. In this specific example, we need some kind of input from the player to determine if they are enjoying the game, either as measures of their experience [247], or their selection of difficulty settings. This input that infers the emotional state of the player and can be used for adaptation and successful emotion elicitation is referred to as the emotional input.

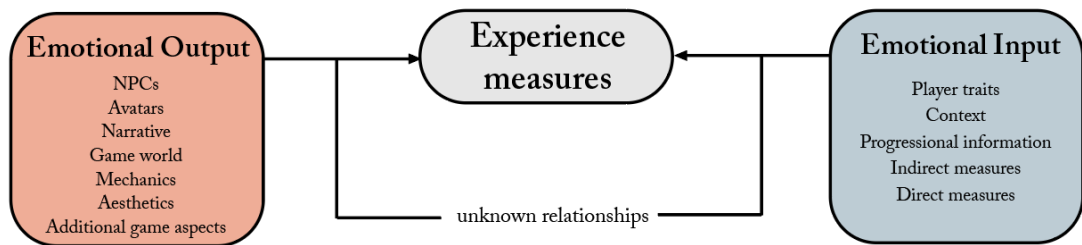


Figure 4.3: Model of relevant variables for the user research that allows the connection of emotional output and emotional input through experience measures given a game context. The tested statistical relationships can be used to inform the design of emotional evocative games, without the need for specified emotion models.

4.2.3 Second domain: Integrate Emotional Input

4.2.3.1 Identify of fitting measures

Emotional input in a game design context could be any data that holds information about the emotional player profile in the current moment of the game-player interaction. This includes information about players and game contexts that have been established in the previous steps, but also other measures of affect. For example, if user research identified that the target experience of melancholy was reliably elicited via a fixed narrative, developers can reasonably assume that the players experience melancholy when playing through the events of the game. Emotional input can however include additional direct and indirect

measures of emotions:

The most prevalent measures are of physiological nature and multiple instruments are commonly investigated: electrocardiogram (ECG) measures heart activity, which can include both heart rate (HR) and heart rate variability (HRV). In gaming contexts, HR and HRV are often related to stress or fear [44], [215], [229], [248]. The main benefits include low cost and straightforward interpretation of information, but measures can only make predictions about arousal. Similarly, Electrodermal activity (EDA) or Galvanic Skin Response (GSR) refers to the skin's electrical conductivity, which can be measured through sensors that are small, wearable, and cheap. It is commonly used to reflect arousal aspects of emotions, which is used in some affective gaming studies to infer information about high-arousal emotions, such as fear or excitement [207], [248], [249], but has limits in making distinctions between these emotions. Electromyograph (EMG) measures the electrical potential of muscle cells, which is commonly used in gaming contexts to make inferences about valence aspects of emotions by measuring facial muscle activities [207], [250]. While this is a promising approach to make inferences about distinct types of emotions, EMG facial sensors can be more invasive than ECG or EDA sensors and interpretations are reliant on the coding system that relates muscle activity to emotional states, which potentially limits its use for a range of more complex emotions [251]. Finally, Electroencephalography (EEG) is used to measure the electric activity of the brain, which has been found to be closely related to reports of valence and arousal but includes multiple, invasive sensors and complex data that is sensitive to any changes in brain activity [145] and is therefore difficult to associate with emotional responses. Whether or not specific emotions map onto specific physiological profiles across all contexts remains unclear, so relying upon a given physiological measure or group of measures to reveal a given emotion is conceptually problematic. Nevertheless, physiological sensors offer a wide range of primary affective data that can be used to test assumptions about their relationships with target emotional experiences [32], [252].

Another class of measures considers observational behaviours, such as gestures, postures, facial expressions, or eye gaze. Motion captures have had some success in classifying emotions in regards to either high or low intensity [253]. Measurement of facial expressions has been used to distinguish between emotional facial features and is reported to be easy and unobtrusive to implement [166]. However, accuracy seems to depend on players being in the right position, being similarly expressive, and not manipulating reactions [44].

Conflicting evidence also suggests that certain emotions can be expressed in multiple ways and are not universally reflected across cultures [254], which could limit the practicability of such an approach. Another measure lies in eye gaze, which is typically associated with attention, but has also been used to provide simple classification information on arousal or valence aspects of emotions [255], [256] or to create new gameplay input systems [257]. While eye-gaze is hard to track in classical gaming setups, virtual reality headsets with included eye-tracking devices are becoming more popular, which makes this technology for some games easy to implement and a promising source of data for adaptation [258], [259].

Finally, in-game behaviour can be used to infer information about current emotions as players have the opportunity to influence the game through their input (or player choice [131]). For example, input behaviour in mobile games has been shown to predict low and high levels of valence and arousal [260] and certain emotional states are related to input variables like touch pressure [230] or controller button presses [261]. There have been efforts made to model emotional states from in-game performance or input parameters (e.g. [231], [262]) and most recently, emotion classification based on in-game responses to narratives has been used to measure distinct emotional states [263]. This is still a young field, but game adaptation based on behaviour yields promising results in engaging different types of players [232], [264].

In summary, there are many possible sources of information, although none are universally applicable. The choice depends on the design aim and requirements, but should in every case be informed by user research.

4.2.3.2 Examine relationships between users and measures

As with the relationship between game material and emotional reaction, we often cannot assume that the relationship between player input and their current emotional state is known. Although there is a wide variety of measures that have been related to emotion in one way or another, there is no perfect solution to measure emotions in games. Robust, universal one-to-one matchings between emotions and measurement either have yet to be discovered or do not exist. Some researchers in the field argue for a combination of different methods, which seems necessary to increase accuracy, but also increases cost and effort.

Designers can however avoid these uncertainties by focusing on the project's specific design aims. Since the resulting subjective experience is the design aim, games do not need to accurately predict all emotions, but rather explore the relationship between appropriate

available data and the subjective experience of the target emotion. When testing the effects of the games on user's emotions through user research (as explained in Section 4.2.2.2), games can also test what kinds of user input can be used to assess the current player's emotions.

This means that the data should provide some indication of the relative presence or absence of the targeted emotional state within the tested game context and for the tested audience. However, it need not distinguish between that state and other unlikely states. A game that aims at relaxation does not need to rely on accurately measuring sadness. However, it should establish some kind of indication of subjective relaxation.

In order to be considered adaptive the game must demonstrate an effect on player relaxation for different individuals and during different phases of the game. The main question to address in the user research step for the second domain is therefore: What information can a game use to identify such differences and adapt? For a physiological relaxation game that uses HR measures user research can explicitly test the relationship between HR and relaxation for different individuals. If it exists in the context of the game and is usable, the game has now a way to track the target emotion of the player and adapt to ensure it, regardless of the individual differences or the current point of the game. For another relaxation game without physiology tracking, user research can for example explicitly test the relationship between in-game performance and relaxation and if this proves useful can then design difficulty levels that ensure a specific in-game performance and therefore relaxation.

Like in the first domain, measures can become usable to infer emotional states by specifically testing their statistical relationships within the context of the game. An illustration of the user research for both emotional output and emotional input is depicted in **Fig 4.3**.

4.2.3.3 Operationalise measures

User research informs about the data sources that can be used to make inferences about emotional states, so the last remaining step is to operationalise these data sources and connect them with the game output, as depicted in **Fig 4.3**. Both the emotional input and the emotional output of a game have an established statistical relationship with the target experiences, so the input can be used to determine the output.

An illustration of such a process can be seen in the study by Shaker et al. [265] who collected gameplay behaviour and self-reported experiences from players first and then used

the findings to identify important input patterns that reflected frustration, challenge, and fun. Game worlds were then created that adapted to these input patterns, which resulted in a better evaluation of the game. Consequent studies showed that collecting gameplay data beforehand provided the opportunity to model player experiences, which could point to the practicability of such a process for almost all games [231].

In this sense, the emotional input and the emotional output are both representations of the experience and are therefore easily exchangeable and enhanceable in reaction to the needs, technical progress, and also to more informed theories. Such a modularization is currently heavily emphasized by researchers in affective computing, not only because it can make actual solutions easier and more approachable, but also because theories from psychology, design, and HCI are more and more in the process of unification and a system that draws and informs all of these fields will need to be dynamic [266]. For example, if face recognition is used as a measure for anxiety within a game, it can be complemented by other measures or even easily exchanged, because it only represents the tested relationship with the experience and not an underlying model of emotion that makes face recognition necessary.

4.2.4 Summary

To ensure that a game elicits a target experience, this framework provides the terminology and methodology to make some of the unknowns explicitly known and usable to design game elements. This process includes:

First Domain: Design Emotional Output

- 1. Identify fitting stimuli.** Make use of common design practices (e.g. the use of successful examples, and game design frameworks) to identify game elements with the potential to elicit the target experience.
- 2. Examine the reaction of specific users to stimuli.** Explicitly test the relationship between the designed game characteristic and the target emotion using user research methods. This step needs to take individual differences and game progression into account.
- 3. Design stimuli and rules of adaptation.** Given the knowledge of step 2, determine how the game needs to adapt between individuals and during the progression of the game to ensure the target experience. This step needs to take the second domain into account.

Second Domain: Integrate Emotional Input

1. **Identify fitting measures.** Identify ways to indicate whether or not the target emotion is experienced with consideration to cost, invasiveness, generalizability, and ease of interpretation.
2. **Examine the relationship between users and measures.** Explicitly test the relationship between the chosen measures and the target emotion using user research methods.
3. **Operationalise measures.** Given the knowledge of step 2, determine how user input can be used to trigger the adaptations of the first domain.

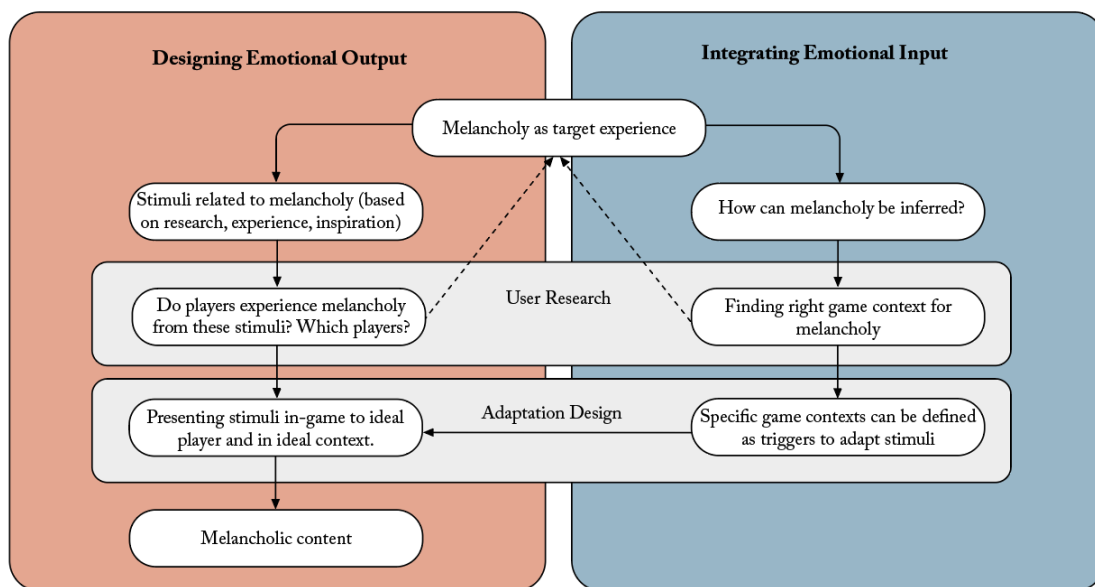


Figure 4.4: Example of process application using melancholy as an aesthetic experience.

This framework was developed to help avoid outdated classification techniques or simplified theoretical assumptions. To bring video games closer to the benefits associated with adaptation, this process can be followed for certain games or game aspects, while issues currently present in affective games research (as such rules that are unlikely to hold across all situations and individuals) are being addressed. Designers and developers should be able to test relationships themselves for a specific game context and player group to make informed decisions about how to adapt a game to player emotions. This is necessary because there are many unanswered questions about the nature, variety, and underlying mechanisms of emotion. This framework does not claim to provide answers to these ques-

tions; instead, the aim is to illustrate methods that could help designers and researchers get closer to finding some of the answers themselves. In fact, because we are far from a common understanding of how emotions can be modelled, it is more useful to enable developers and researchers to contribute without the need for a comprehensive understanding of emotions.

4.3 Conclusion

To address the problems present in affective games, it is important to synthesize our current knowledge in our psychological emotion research, design research, and HCI research with a focus on affective computing. This chapter collates relevant research findings to provide a model of the ongoing emotional interaction between a player and a game that can be represented by our refined version of the emotional feedback loop (**Fig 4.1**). Based on this loop, a methodological framework of emotion design for video games was illustrated that introduced explicit methods to implement emotion-adaptivity in games and to research some of the open questions that remain about our emotional connection to games. As demonstrated in Chapter 3 and mirrored by research over the years [267], there is a lack of unity in the work that conceptualizes game-player interactions. This approach takes a step towards reducing this lack. It connects theoretical perspectives from multiple disciplines. The proposed approach was created to reduce confusion and to provide guidance for applied emotion research games that do not restrict them by oversimplifying their assumptions about emotions or misinterpreting relationships between affective variables. It is crucial to not expand the confusion even more but to connect different fields with a common theory that is both dynamic and useful. Doing this, a design-led approach could address some of the technological and conceptual problems that currently exist in the affective game literature. An emotional game is not (yet) a game that completely understands the full range of human emotions, but a game that can successfully create an emotional experience. Only with an approach that is robust to uncertainties, but still enabling in the questions that matter, can designers and researchers understand more about the processes that are involved.

To achieve this, it is necessary to examine how this framework could be integrated with common design practice in a way that supports the process. An important step to demonstrate the usefulness of this framework is a thorough evaluation of the process as a whole and of the concepts introduced here.

Chapter 5

Affect in Adaptive Systems: Creating "The Flow Experience"

*"To this state we have given the name of "flow",
using a term that many respondents used in their
interviews to explain what the optimal experience
felt like."*

— Mihaly Csikszentmihalyi [133]

With a better grasp on the problems in the field of affect-adaptive games (Chapter 3) and a potential process towards mitigating many of these problems with the Emotion Design Framework (Chapter 4), the logical next step is to put the framework to the test. In order to achieve this, this chapter will follow the structure shown in Fig 5.1: First, emotional input and emotional output variables will be identified for a newly developed video game. Following this, two empirical experiments will be constructed, one mirroring the testing phase of the Emotion Design Framework, and the other testing the result of the adaptation phase against a non-adapted game. To be more specific, in the first experiment, a video game will be developed with a set of potential emotional output variables and potential emotional input variables. The statistical relationship between these variables and the subjective emotional experience of the players will be tested. Based on the results of these tests, the second experiment - the adaptation phase - will implement affective game adaptation and test it against a control version of the same game with rigorous empirical standards. In doing so, this chapter aims towards providing validation for the work given in the previous chapters and expand on the findings with data-driven evidence.

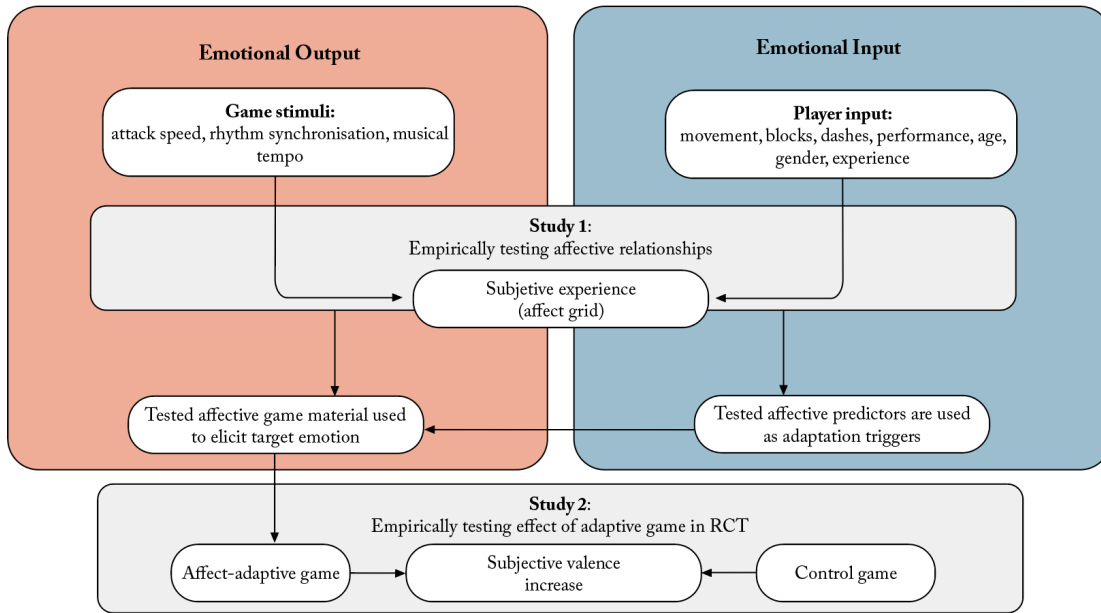


Figure 5.1: Illustration of the structure of this chapter: Each empirical experiment targets a specific process in the emotion design framework. Study 1 showcases the user research phase and the resulting information is used to build an affect-adaptive game, which is evaluated in Study 2 against a control game.

5.1 Aims and Motivation

The main aim of this chapter is to expand on the typical process of developing affective games by focusing on explicit testing of emotional associations between game data and subjective player experience before the adaptation is integrated into game development. This means that before an adaptation is proposed and planned, user research for a newly developed game is conducted to find potential affective associations between emotional output, emotional input, and subjective player affect. The process reflects the **testing phase** and is illustrated in Fig 5.2.

Following this, affective adaptation processes are designed and built based on the gathered data from the testing phase. These are then in turn empirically evaluated by testing the effect of the adaptive game against a control version of the game in terms of increased game enjoyment. The game designed for this study was "The Flow Experience" a 2D top-down arcade-style game with the specific aim to evade enemy attacks for a set amount of time.

The first experiment involves exploratory analyses of in-game variables to test which

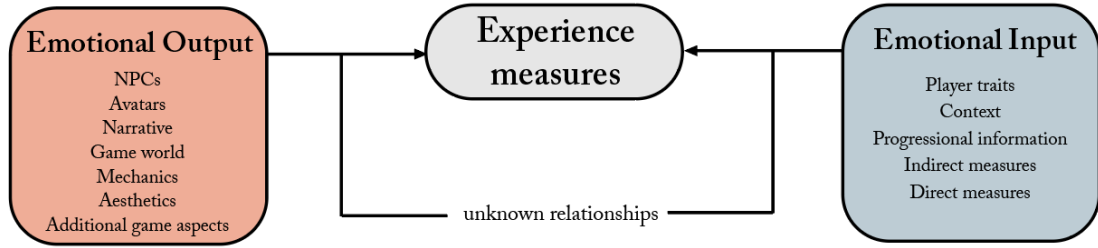


Figure 5.2: Illustration of the relevant emotional information to design and test affective video games. We describe information that could infer the emotional state of the player as potential emotional input. Information that could influence the emotional state of the player is described as potential emotional output. In order to create a functioning affective loop, the relationship between subjective experience and both potential emotional input and output must be examined.

could be used to indicate the target emotion, which is defined as the change in experienced valence (i.e. the ability of the game to increase positively-valenced experiences), which was chosen as a valid affective construct to avoid conceptual uncertainties of game enjoyment measures [268], as well as experienced arousal, which is widely used in combination with valence to describe affective experiences [32]. It is important to note that we use valence and arousal as a dimensional construct to purely measure subjective affect and not as a definition of emotional state, therefore avoiding fundamental assumptions. Confirmatory analyses are then used to test the success of specific game mechanisms in eliciting the target emotional state. These mechanisms were chosen because of their theoretical ability to influence experienced valence. Based on the research of dynamic difficulty adjustment (DDA) that provides a body of work showcasing a relationship between challenge and enjoyment [269], enemy attack speed was considered as a candidate variable to manipulate. It was hypothesized that attack speed would influence the emotional experience for certain types of players (i.e. players that enjoy easier versus harder games) and that these players could be identified by measuring in-game data. Based on research showcasing the emotion-evocative power of adaptive music [270]–[272], both tempo and synchronization of in-game material to the beat were considered as candidate variables. It was hypothesized that slower music would contribute to relaxing experiences and that synchronization between the gameplay and music would lead to a valence increase. Findings from the first experiment

were then used to create adaptive versions of the game.

In the second experiment, these adaptive versions were tested against a control group to evaluate the whole process. The main aim was to provide an example of the novel methodological process presented in the previous chapters that can be used to test emotional assumptions explicitly, based solely on subjective affect measures and use this information to make a more enjoyable adaptive version of the test game. As previously presented, this process was designed to be robust against contentious theoretical assumptions and make different affective games more comparable, which ultimately could help provide more unified insights about the effects of emotion adaptation in video games and how emotions function in real-life applications.

5.2 The Game: *The Flow Experience*

For the purpose of this chapter, a modern 2D action game was developed, utilizing arcade-style gameplay, which has been tested in the context of affective adaptivity with mixed results [194], [195]. As reported by Gundry and Deterding [273], there are many potential validity threats associated with experimental game research that need to be addressed, such as games' complexity, sources of unwanted variance, and different social framing. These barriers were aimed to be addressed through the methodological approach. "The Flow Experience" as a game was used to provide an example with high ecological validity, i.e. a game that represents an example of a commercial entertainment game. While it is often argued that commercially available games with an established player base may achieve this the best (e.g. [274], [275]), a new game was created to provide insights about the application of the proposed principles within the design process and not post-hoc for a finished product. Additionally, validity concerns relating to the complexity of game stimuli, familiarity with the test material, and the influence of unwanted variance were more controllable in a prototypical version of a new game that allowed targeted manipulation of variables to make clear causal inferences (see Chapter 5.3.1.1 for more details). The game was developed in the Unity framework, using C# as a scripting language. The completed version of the game is currently available on Steam [5]. A detailed design rationale can be found in B.

5.2.1 Gameplay

The gameplay loop is based on classic arcade games, such as Asteroids. Players are placed in a fixed-sized 2D arena with a top-down view. One enemy moves and teleports around the arena while shooting (see Fig 5.3). To create a very focused gameplay interaction, players' only goal is to avoid getting hit for a level-dependent amount of time. Players have three different mechanics that they can utilize to avoid projectiles: (a) Continuous movement in a two-dimensional space; (b) a dash, which gives a short boost and makes players invincible, but leaves them standing still for a short time afterwards; (c) a block that guards against certain attacks with a shield that has a fixed cooldown time after usage and that prevents further movement. Players start each level with three lives and restart if all lives are depleted. If one life remains for a set amount of time, one wins the level.

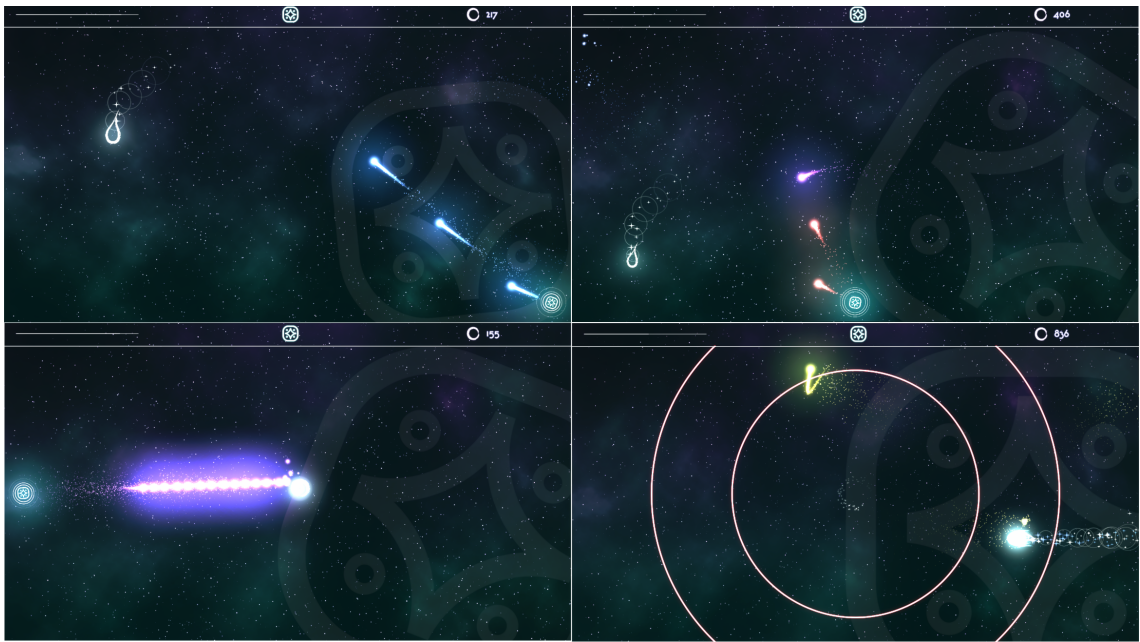


Figure 5.3: Screenshots of the first (top left), second (top right and bottom left) and last (bottom right) experimental level of *The Flow Experience*. Top: The player evades projectile attacks by the enemy through simple movement. Bottom Left: The player blocks a row of incoming projectiles with a shield. Bottom Right: The player dashes through wave attacks to not get hit. The top bar in each screenshot indicates the current state of the level, i.e. the time remaining.

5.2.2 Visual and Sound Design

The game's visual elements were created with two aims in mind: (a) provide a fitting and thematic setting, related to positive experiences as described by Csikszentmihalyi [133], and (b) keep the overall aesthetic deliberately abstract to minimize expectation effects and to limit narrative interpretation outside the scope of the work (see Fig 5.3). To achieve these aims, the setting was chosen to be an abstract representation of space and the player and enemy were represented through ancient symbols. The fairly sparsely orchestrated electronic score was chosen to provide a soft and slow support for the openness of the general aesthetics in order to not contribute to stress, but rather calmness, even in difficult scenarios. In the top-right corner of the screen was a score indicator. For the experimental version of the game, this score did not reflect performance, but rather it increased continuously for all participants and was therefore purely aesthetic.

5.2.3 Pilot Study

A pilot study was conducted to test the experimental version of the game to make it as usable and easy to understand as possible. 10 participants (7 female) with a mean age of 26 (ranging from 18 to 29) were recruited for the pilot study and asked to play through the experimental version of the game, consisting of one tutorial level and three experimental levels. Afterwards, each participant was asked to report any problems or sources of confusion regarding the experimental setup, the tasks, and the game. Based on the reports, some wordings in the affect measurement screen were changed to make it more clear how arousal and valence were defined. Additionally, text was included at the beginning of the game to make players aware of the possibility of remapping controls (both for gamepad and keyboard controls) to their convenience and to improve accessibility.

5.3 Study: Testing Phase

5.3.1 Materials and Methods

5.3.1.1 Measures

The aim of this first experiment is to test statistical relationships between subjective experience and potential emotional output, as well as potential emotional input, as reflected in Chapter 5.2. To achieve this, multiple variables were measured within the game:

For self-reported affect, valence and arousal levels of participants were chosen [61]. Specifically, an affect grid [276] was used, because it is an established, fast, and easily understandable single-item instrument, with high and well-documented validity and reliability for the subjective component of emotions [29]. In theory, any validated self-report instrument could be used at this stage, but the affect grid was chosen, because it provides continuous and well-defined data for subjective affect. The grid has two dimensions, one for arousal (on a 7-point scale) and one for valence (on a 7-point scale).

Several behavioural measures were used to assess potential emotional input, i.e. player input and demographic variables that were then tested to assess how well they represent subjective affect. Each variable was measured within one level, resulting in four distinct sampling points for each variable. These variables included: (a) the number of deaths in a level, (b) time spent moving the character, (c) the number of performed blocks, (d) the number of performed dashes, (e) the number of avoided attacks, (f) how close actions were performed to a downbeat (i.e. rhythmic behaviour; operationalized as the average time difference to the nearest downbeat), (g) lost lives in a level (death and restart occurs after all three lives are lost), (h) number of wins (i.e. number of levels successfully finished, as a level ends automatically after 5 minutes or 6 deaths). Additionally, demographic measures of age, gender, and how regularly participants play games ("Never", "Monthly", "Every two weeks", "Weekly", "Every few days", "Daily") were included.

5.3.1.2 Manipulations

Because this first experiment aims to investigate associations to then inform the design of an adaptive version for the second experiment, four in-game variables were manipulated to specifically test their effects on subjective player affect (representing potential emotional output). The first one was the played levels, which represent differences in in-game contexts between participants. Practically the level determined what enemy patterns players encountered and how much time between different enemy actions (either movement or attacks) passed. To mimic a real gameplay environment, the level order was not randomized but designed to increase gradually in difficulty and specific gameplay aspects (see 5.3.

Three additional independent variables were experimentally manipulated to test their direct effect on player emotions (i.e. how well they influence player emotions). These variables were within-subject manipulations, meaning that their manipulation took place within a player's session and changed between the three experimental levels of the game.



Figure 5.4: In-game implementation of the Affect Grid adapted from Russel [276]

To address sequencing effects, the presentation sequence was counterbalanced between participants. The first one was musical tempo, which was presented in three fixed expressions (120bpm, 140bpm, 160bpm) by either slowing down or accelerating the game's background music. The second variable was a multiplier of enemy attack speeds in three expressions (0.5x; 1x; 1.5x). The last variable was enemy synchronisation, consisting of two expressions. The synced version forced enemy actions always on the downbeat of the music, while the non-synced version used a fixed amount of time (based on the average of the synced version) between enemy actions, ignoring the background music completely.

5.3.1.3 Procedure

The experiment was finalized and a WebGL version of the game was made available online via the free video game hosting platform itch.io. After starting the game, participants were asked to carefully read the study information sheet and agree to voluntarily participate in

the study via a consent form. Participants were then led to the home screen of the game, where they were able to revise and optionally remap the game's control scheme. Following this, participants played through the tutorial level of the game that was designed to teach the three main mechanics of the game: Moving, blocking, and dashing (in this order), explaining the controls again throughout.

Players then played through the three experimental levels that were presented in a fixed order as a way to mimic natural progression in video games. Before and after each level, players were led to the affective grid screen to indicate their levels of arousal and valence. Each experimental level was designed to include enemy patterns that revolve around one of the three main mechanics: Moving in the first level, blocking in the second level, and dashing in the third level. Every participant encountered manipulations of one of the three manipulated independent game variables (either musical tempo, enemy attack speed, or beat synchronization), while the other two variables were kept stable. Manipulations of each variable changed per level and the presentation sequence was counter-balanced between participants.

5.3.1.4 Participants and Statistical Analysis

A total of 452 participants were recruited through social media (Reddit, Twitter, Facebook) by posting the link to the study in appropriate groups and sub-forums, as well as through the institutional subject pool of the authors. Participation was voluntary and compensated through University credits if applicable. To minimize the risk of error due to participants not taking the experiment seriously, participants who withdrew at any point during the experiment were excluded, leading to a final sample size of 161 (88 female). The sample size was considered appropriate based on a statistical power analysis, yielding a power of 0.95 for medium-sized effects (0.5 SD) and 0.8 for small effects (0.25 SD) in repeated measures ANOVAs. Participants' age ranged between 18 and 45 years ($M=22.04$; $SD=6.30$). 19% reported never playing games, 20% reported playing monthly, 6% reported playing biweekly, 9% reported playing weekly, 19% reported playing every few days, and 27% reported playing daily. After removing incomplete cases and computing demographic data, three steps of statistical analyses were conducted. First, an exploratory analysis utilizing correlation and multiple regression analysis was used to detect associations between player behaviour variables (potential emotional input) and affective measures. Regressions included interaction terms and were used to test the best fit of subjective affect predic-

tion. Second, a confirmatory analysis was used to statistically test the effect of in-game manipulations (potential emotional output) on affective variables. This was done using a Within-Subject ANOVA with follow-up t-tests to gather more information about observed effects. Because only one of the three emotional output variables was manipulated for a given participant, these analyses were done separately for three groups: One for musical tempo (N=55), one for attack speed (N=57), and one for enemy synchronisation (N=49). Third, to test if any observed effects in Step 2 were dependent on player or game variables identified in Step 1, regression analysis was performed using F-tests and follow-up t-tests. This was done to identify interactions between emotional input and emotional output in predicting user experience as a way to propose potential adaptation mechanisms. All analyses were conducted in R.

5.3.1.5 Ethics Statement

Written consent was granted after reviewing the methods of our study by both the Physical Sciences Ethics Committee and the Ethics Committee of the Psychology Department of the authors' institution. The experiment was conducted in accordance with the recommendations of these committees.

5.3.2 Results

5.3.2.1 Explorative Analyses

Changes in arousal and valence ratings were calculated by subtracting measures made after a given level from measures made before a given level. Descriptive data for mean changes of arousal and valence across all levels can be seen in Table 5.1. To measure the relationships between changes in emotion ratings and game data, correlations were calculated for each game variable with arousal change and valence change. Results showed multiple statistically significant associations. Changes in arousal were positively related to the time spent moving (*Pearson's* $r = .23$; $p = .003$) and to the number of avoided attacks (*Pearson's* $r = .20$; $p = .01$). Changes in arousal were also negatively associated with the number of wins (*Pearson's* $r = -.18$; $p = .02$)

Changes in valence were positively related to the number of blocks used by players (*Pearson's* $r = .24$; $p = .002$), as well as with the number of avoided attacks (*Pearson's* $r = .22$; $p = .004$;). Changes in valence were also negatively associated with the number of deaths in a level (*Pearson's* $r = -.25$; $p = .002$;), and with the number of lost lives within

a level (*Pearson's* $r = -.25$; $p = .002$). No other correlation was significant.

Group	Valence Change			Arousal Change	
	N	M	SD	M	SD
Low attack speed	57	0.09	1.62	0.05	1.30
Medium attack speed	57	-0.40	1.52	0.30	1.51
High attack speed	57	-1.11	1.13	0.74	1.43
Low tempo	55	-0.23	1.96	0.13	1.31
Medium tempo	55	-0.02	1.81	0.24	1.43
High tempo	55	0.35	2.17	0.35	1.44
No synchronization	49	-0.82	1.68	-0.51	1.70
Synchronization	49	0.14	1.06	0.06	1.34

Table 5.1: Descriptive data over all three experimental levels by conditions.

Using these correlations as a quality measure for the prediction, regression models were tested that could explain the most variance of arousal and valence. Multiple regression models were tested against each other, using all possible combinations of variables showing a significant association with affect. The respective model with the highest R^2 for valence and arousal was deemed as the model with the best fit. For arousal, a linear model using movement (*standardized* $\beta = 0.23$ $p < .001$) and number of wins (*standardized* $\beta = -0.18$; $p = .02$) showed the highest effect ($F[2, 159] = 7.32$; $p < .001$; $R^2 = .08$). For valence, a linear model using number of blocks (*standardized* $\beta = 0.16$; $p = .06$) and number of deaths per level (*standardized* $\beta = -0.17$; $p = .05$) showed the highest effect ($F[2, 159] = 7.00$; $p < .001$; $R^2 = .08$). These predictors were further used to test interactions on the effects of emotional outcome variables.

To test for influences of individual traits, the associations between demographic data and emotions were examined. Game experience showed a significant correlation with valence ($t[160] = 2.53$; $p = .01$; *Pearson's* $r = .20$), no significant association was found for age or gender.

To test for influences of game context, differences of emotion changes for the different levels were tested via Within-Subject ANOVAs. There were significant effects on both valence ($F[3, 483] = 6.08$; $p < .001$; $\eta_p^2 = .03$) and arousal ($F[3, 460] = 3.70$; $p = .014$; $\eta_p^2 = .02$). Follow-up t-tests revealed that all three levels had a lower change in valence compared to the tutorial level (Tutorial: $M = -0.04$; $SD = 1.62$; Level 1: $M = -0.79$; SD

= 1.57; $t[161] = -4.18$; $p < .001$; *Cohen's d* = -.47; Level 2: $M = -0.43$; $SD = 1.56$; $t[161] = -2.28$; $p = .02$; *Cohen's d* = -.24; Level 3: $M = -0.44$; $SD = 1.56$; $t[161] = -2.38$; $p = .02$; *Cohen's d* = -.25). Similarly, the tutorial level had a higher change in arousal ($M = 0.65$; $SD = 1.32$) compared to the experimental level 2 ($M = 0.23$; $SD = 1.31$; $t[161] = 2.96$; $p = .004$; *Cohen's d* = .32) and 3 ($M = 0.16$; $SD = 1.59$; $t[161] = 2.92$; $p = .004$; *Cohen's d* = .33).

5.3.2.2 Confirmatory Analyses

The effects of the emotional outcome variables (tempo, attack speed, and enemy synchronisation) were tested via Within-Subject ANOVAs. Musical tempo showed no effect on valence ($F[2, 108] = 1.09$; $p = .34$) and no effect on arousal ($F[2, 108] = 0.29$; $p = .74$).

Attack speed showed a significant effect on valence ($F[2, 112] = 8.51$; $p < .001$; $\eta_p^2 = .11$), and a borderline effect on arousal ($F[2, 112] = 3.04$; $p = .052$; *Eta Sq* = .04). Follow-up t-tests revealed significant effects on valence for low attack speed ($M = .09$; $SD = 1.62$) vs. high attack speed ($M = -1.12$; $SD = 1.13$; $t[56] = -4.38$; $p < .001$; *Cohen's d* = -.58) and for medium attack speed ($M = -.41$; $SD = 1.52$) vs. high attack speed ($M = -1.12$; $SD = 1.13$; $t[56] = -2.96$; $p = .005$; *Cohen's d* = -.39).

Enemy synchronization showed no effect on valence ($t[48] = 1.03$; $p = .31$) and also no effect on arousal ($t[48] = -0.31$; $p = .76$).

5.3.2.3 Interaction Analyses

To test whether the effect of attack speed on emotion ratings was dependent on the identified emotional input variables, regression analyses were conducted for each of the following predictors: (a) game experience; (b) the best-fit model of player data, using number of blocks and number of deaths per level; (c) game context (i.e. played level). Each regression predicted emotion ratings from one of these variables, speed, and the interaction between the two. To make interaction analyses possible, these emotional input variables were transformed into three groups (low, medium, and high), based on the 33rd, 67th, and 100th percentile. No interaction was found for game experience ($F[2, 112] = 1.65$; $p = .18$), but the linear model using number of blocks and number of deaths per level showed an interaction effect ($F[4, 92] = 3.41$; $p = .02$). Posthoc-tests revealed that for the low player model group, low attack speed led to higher valence increase ($M = 0.21$; $SD = 1.47$) compared to medium ($M = -0.84$; $SD = 1.34$; $p = .02$) and high attack speed ($M = -0.79$;

$SD = 1.08$; $p = .02$). Similarly, for the medium player model group, the low attack speed led to higher valence increase ($M = 0.16$; $SD = 1.34$) compared to medium ($M = -1.00$; $SD = 1.05$; $p = .004$) and high attack speed ($M = -1.26$; $SD = 1.19$; $p < .001$). However, for the high player model group, the medium attack speed led to higher valence increase ($M = 0.63$; $SD = 1.61$) compared to low ($M = -0.10$; $SD = 2.02$; $p = .03$) and high attack speed ($M = -1.26$; $SD = 1.10$; $p < .001$).

Player level also showed a significant interaction effect ($F[4, 162] = 2.59$; $p = .04$), as illustrated in Fig 4. For level 1, low attack speed led to a higher valence increase ($M = 0.05$; $SD = 1.72$) compared to medium attack speed ($M = -0.84$; $SD = 1.34$; $p = .009$). For level 2, the high attack speed showed a significantly smaller valence increase ($M = -1.33$; $SD = 1.28$) compared to low ($M = 0.31$; $SD = 1.08$; $p < .001$) and medium attack speed ($M = 0.30$; $SD = 1.69$; $p < .001$). For level 3, the low attack speed led to a significantly higher valence increase ($M = -0.05$; $SD = 1.90$) than the high attack speed ($M = -1.15$; $SD = 1.18$; $p = .01$). Interactions are illustrated in Fig 5.5.

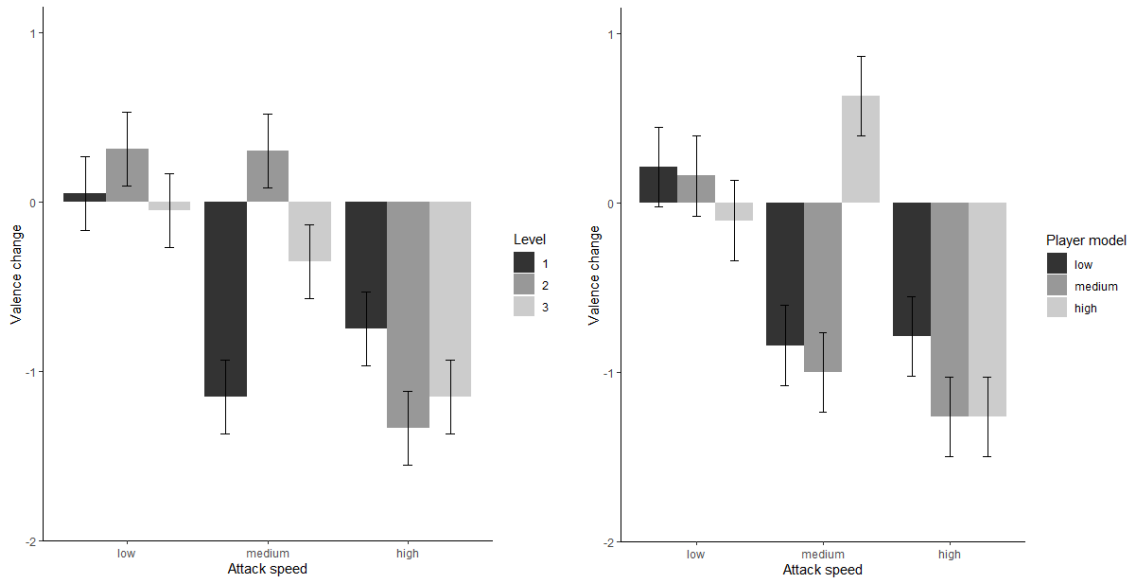


Figure 5.5: Bar graphs of the interactions. Left: Effect of attack speed on valence change by level (game context). Right: Effect of attack speed on valence change by expressions of the linear player model consisting of the number of blocks and number of deaths. Error bars symbolize standard error (SE).

5.4 Study: Adaptation Phase

Based on the results of Experiment 1, two potential adaptation mechanisms were identified. The first one is game level, which showed a significant interaction with the effect of attack speed on valence. More precisely, as Figure 5.5 shows, the experimental levels 1 and 3 showed that low attack speed was enjoyed the most, while in level 2 low and medium attack speeds were enjoyed equivalently. The second one was the linear model using the number of blocks and number of deaths per level, which could be seen as a proxy for player skill in this specific game scenario. As Figure 5.5 shows, players with a low or medium expression in this model had the biggest valence change with the low attack speed, while players with a high expression in this model had the biggest valence change with the medium attack speed.

Based on these results, two types of adaptation are proposed: (a) Level-based adaptation that has a medium attack speed in level 2 and a low attack speed in levels 1 and 3; (b) Player-based adaptation that dynamically changes attack speed based on a linear regression model identified in Experiment 1 (utilizing number of blocks and number of deaths of players), with low attack speed for players who perform same or lower than the 67th percentile of participants from Experiment 1 and medium attack speed for players who perform better than the 67th percentile of participants from Experiment 1. The following experiment aims to evaluate both of the proposed adaptations by comparing them to a control group with no adaptation and a fixed medium attack speed.

5.4.1 Materials and Methods

5.4.1.1 Design

The experimental setup was the same as in Experiment 1, using The Flow Experience as a test game with the same progression (tutorial level and same three experimental levels). This time, musical tempo (set to 1x speed) and beat synchronization (turned on) were not manipulated, only attack speed was manipulated between players. Attack speed was manipulated differently based on three conditions: (a) the control condition was the same as in Experiment 1, using a 1.0 multiplier for attack speed in every situation; (b) for the context-based adaptation attack speed was multiplied by 0.5 in level 1 and 3 and by 1.0 in level 2; (c) for the player-based adaptation attack speed multiplier was based on a linear regression formula which was clamped between 0.5 and 1.5, dynamically changing

the difficulty through gameplay for every attack:

$$v = v_i * \max(0.5, \min(-0.34 + (0.034 * n_b) - (0.117 * n_d), 1.5))$$

with v as the attack speed, v_i as the original attack speed, n_b as the number of blocks, n_d as the number of deaths. Participants were randomly assigned to one of the three conditions and had no knowledge of the existence of other conditions or the nature of their condition. Again, valence and arousal were measured as dependent variables through the affect grid [61].

The general procedure was the same as in Experiment 1, with the same online setup using itch.io. First, participants were asked to read an information form and give consent to participation. They were able to review and modify the game controls and then start the tutorial level that was used to explain and teach the mechanics. Players then played through the same progression of the three experimental levels and measured their affective valence and arousal through self-reporting before and after each level. Every participant played through one of the three conditions (context-based adaptation, player-based adaptation, control), based on randomization.

5.4.1.2 Participants and Statistical Analysis

A total of 245 participants participated in Experiment 2, while the recruitment process was kept the same. In order to test both group differences and level effects, participants who withdrew during the study were excluded, reducing the final sample to 158 participants (100 female). The sample size was deemed appropriate following a power analysis, assuming a medium effect size (0.6 SD) as found in Experiment 1 and a target statistical power of 0.95 (minimum sample 120), as well as considering the recommendation by Ferguson et al. [277], with a target statistical power of 0.8 and a smallest effect size of interest of 0.41 SD, which is argued to represent "practically" relevant effects for media data (minimum sample 156). Participant age ranged between 18 and 45 years ($M=23.25$; $SD= 7.69$). 36% reported never playing games, 15% reported playing monthly, 7% reported playing biweekly, 4% reported playing weekly, 13% reported playing every few days, and 25% reported playing daily. After removing incomplete cases, the data were used to first gather demographic information about the sample. Analyses consisted of mixed ANOVAs with changes in valence and arousal as the outcome variables, played experimental level as a within-subjects predictor, and experimental condition as a between-subjects predictor. Follow-up t-tests were used

to better describe differences. All analyses were again conducted in R.

5.4.1.3 Ethics Statement

As with Experiment 1, the experiment was conducted in accordance to the recommendations of the Physical Sciences Ethics Committee and the Ethics Committee of the Psychology Department of the authors' institution.

5.4.2 Results

The ANOVA on valence showed a significant effect of adaptation condition on valence change ($F[2, 155] = 12.38; p < .001; \eta_p^2 = .05$). There were no differences for played level as the within-subjects condition ($F[2, 310] = 0.56; p = .57$) and no interaction between adaptation condition and played level ($F[4, 310] = 1.80; p = .13$).

Condition	Valence Change			Arousal Change	
	N	M	SD	M	SD
Level 1	158	0.10	1.53	0.28	1.46
Level 2	158	-0.07	1.47	0.29	1.59
Level 3	158	-0.01	1.39	0.05	1.63
No adaptation	52	-0.38	1.47	0.11	1.38
Level-based adaptation	57	-0.09	1.53	0.28	1.71
Player-based adaptation	49	0.50	1.32	0.22	1.52

Table 5.2: Descriptive data for each experimental level and adaptation condition.

Follow-up t-tests revealed significantly higher valence for the player-based adaptation ($M = .50; SD = 1.32$) against both the level-based adaptation ($M = -0.09; SD = 1.53; t[223] = 3.54; p < .001$), and the control condition without adaptation ($M = -0.38; SD = 1.47; t[251] = 4.84; p < .001$). Cohen's d for the difference between player-based adaptation and level-based adaptation was larger than the pre-defined smallest effect size of interest (Cohen's $d = 0.43; 95\% CI[0.20, 0.65]$), which was also the case for the difference between player-based adaptation and control condition (Cohen's $d = 0.59; 95\% CI[0.35, 0.83]$). The difference between level-based adaptation and control condition remained non-significant ($t[314] = 1.91; p = .06; Cohen's d = .21$).

There were no effects of adaptation condition on arousal ($F[2, 155] = 0.50; p = .61$), and also no effects of level for arousal ($F[2, 155] = 0.67; p = .51$).

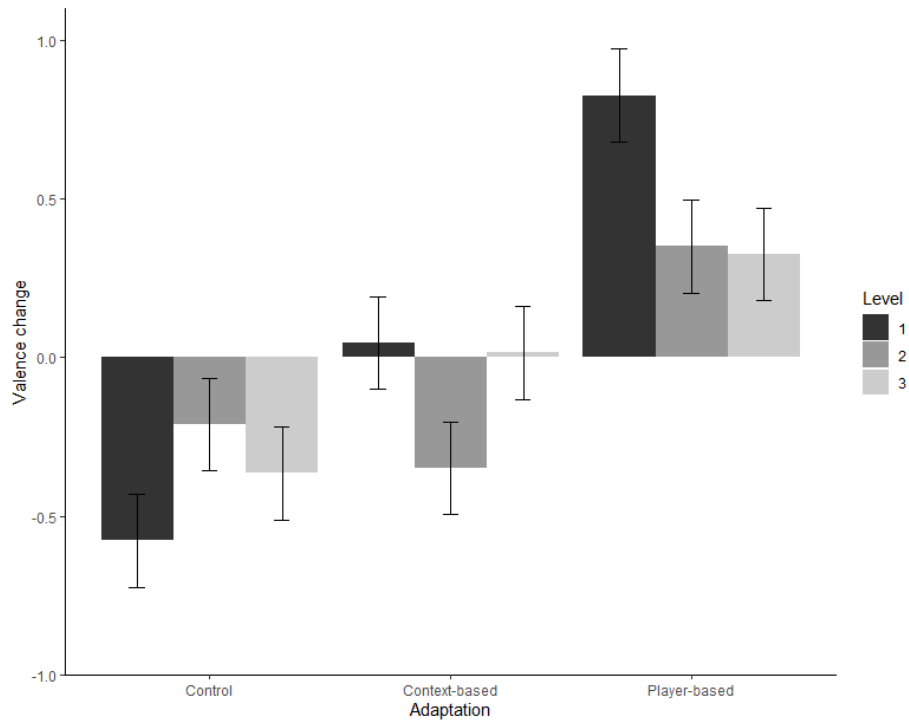


Figure 5.6: Bar graphs of the effects on valence change. Error bars symbolize standard error (SE).

5.5 General Discussion

In Experiment 1, multiple associations between in-game behaviours and emotion were observed within the non-adaptive game, as well as an effect of enemy attack speed manipulation on emotional valence. Based on these findings, two game adaptations manipulating enemy attack speed were designed for Experiment 2: one using in-game behaviours as a way to indicate emotional state and the second using the current level. These were tested against the non-adaptive game version. The first adaptation showed a significantly higher valence increase against both other conditions, providing evidence for the utility of this approach.

5.5.1 Associations Between Game Characteristics and Player Emotion

Experiment 1 used a bespoke 2D arcade-style action game, "The Flow Experience", to test associations between game materials (potential emotional output), player data (potential emotional input) and a self-reported emotion measure. The Flow Experience was found to have an influence on experienced valence and arousal and this influence was in turn at least

partly dependent on certain design decisions (i.e. in-game level) and player abilities (i.e. use of blocks and number of deaths). From a design standpoint, the emotional variables of interest were the "gained valence", i.e. the change in experienced valence before and after playing, as well as the change in experienced arousal. This measure was used to counter conceptually problematic PX outcomes [268] and to serve as a validated psychological instrument to indicate continuous subjective affect [276] (also discussed as the subjective feeling component of emotion in [29]). In fact, while many affect-adaptation studies argue for the use of concepts such as flow [188], [194], [211], more and more arguments emerge indicating doubt of the accuracy relating the concept. For example, Fong et al. [224] collated evidence concerning the relationship between challenge, skill, and motivation and found mixed results, while Jalife et al. [223] provided evidence of inconsistencies between the prevalent concept of flow and current developments in cognitive science. Similar validity and reliability concerns have been reported for other PX constructs, such as enjoyment or immersion [216], [217]. To avoid these concerns, this study provides some evidence of the utility of self-report valence and arousal measures to research player experience. The proposed process can however work with any validated subjective experience measure, as long as it does not make mapping assumptions that are not fully researched (i.e. low physiological arousal and valence mapped to sadness) in order to avoid wrong interpretations.

Using such an approach, the exploratory study yielded multiple relevant insights. Looking at the effect of game levels, we observed a significant decrease in valence and arousal after the tutorial (i.e. the beginning of the experimental levels) and no more main effects between the levels, which indicates the emotion-eliciting effect of the game itself. Both change in valence and change in arousal were related to in-game variables in The Flow Experience, either uniquely (such as movement for arousal or number of blocks for valence), or shared (such as the number of avoided attacks). Valence was more related to game variables that might indicate player skills (such as number of deaths), providing more evidence that game enjoyment is at least in some way related to the subjective experience of perceived challenge, which fits with current views of self-determination theory as an explanation of playing motivation [16]. According to the theory, one relevant basic psychological need that might explain video game motivation is perceived competence, which is dependent on multiple factors. While this study does not measure perceived competence, the found player data indicates that skill-related aspects of the game partly determine valence, which is in line with self-determination theoretical assumptions [19]. Arousal on

the other hand was related to movement, indicating that observable in-game behaviours likely reflect emotional reactions (i.e. game data could indicate affective action tendencies [57]). Arousal also generally decreased after the tutorial level, which fits the notion of affective habituation, making familiar materials generally less arousing [278]. Through the exploratory analysis, it was possible to identify potential variables of interest that are related to valence increase. Such an approach is likely to be useful in exploring affective relationships within games to gather more insights about how emotions can be elicited, measured, or even understood as a psychological construct. The exploratory analyses shown in this study provide just a small glimpse of what is possible to investigate emotions in games and many more aspects could be utilized.

Regarding potential emotional output, the confirmatory analysis provided empirical evidence for the influence of attack speed on valence change. Contradicting some of our hypotheses, musical tempo and beat synchronization showed no effect on emotional reactions. While the means of valence and arousal change as seen in Table 5.1 show the hypothesized trends, there was a large variance found across participants, leading to inconclusive results. Aljanaki et al. [279] conducted a large-scale study about game music ($N = 1778$) and found that individual differences had an especially large influence on the effect of emotion-elicitation by music, which could explain the found results. While this study can not provide specific explanations for the effects, further specific hypotheses can be developed following these findings. From a practical perspective, the lack of an effect could mean that both musical tempo and beat synchronization may not be considered appropriate candidate variables for adaptation, given the specifics of The Flow Experience. In any case, the null results hold no definitive answer, so further studies would be necessary to identify context, users, and manipulation techniques to better explore the effect of tempo and beat synchronization on player emotions. For feasibility reasons, the further design process therefore only regarded attack speed as the manipulation variable of interest.

More insights about the emotional relationship between the player and the game were achieved by investigating the interactions between identified emotional input variables and identified emotional output (attack speed) when explaining observed changes in valence. Two interactions were observed that provided very specific indications of how the game could be adapted to maximize positive valence. The first regarded the game context: In level two the low and medium attack speed were associated with higher valence, compared

to the high attack speed. In levels one and three, only the low attack speed was associated with higher valence. This could indicate differences between the levels in regards to their emotion elicitation and a potential adaptation could change the attack speed from levels one and three to low and to medium for level two. Either way, the results indicated that the high attack speed was too high for every level and that some emotional optimization could be achieved with attack speed balancing in every level. The second interaction regarded player data, Players who showed a high number of blocks and a low number of deaths had the highest valence increase with the medium attack speed, while other players had the highest valence increase with the low attack speed. The relationship between the variables was quantifiable by a linear regression formula that predicted valence change by the number of blocks and deaths. Using this formula, it was possible to differentiate between players and achieve a dynamic difficulty adjustment (DDA) based on emotional data. Affect-based DDA has been shown to benefit player performance [192] and self-reported player experience facets [207], [262], based on a variety of emotion indicators, such as physiology, facial expressions, and self-reports. Because this adaptation emerged as a way to potentially increase valence based on the patterns found in this design process, unknown relationships regarding challenge, skill, and player emotions were not necessary to model to potentially achieve similar effects.

5.5.2 Adaptation Strategies Based On Emotional Information

Experiment 2 provided an evaluation study to test the effect of the two affect-based adaptation mechanisms identified in Experiment 1 against a control group. The player-based adaptation outperformed both the control group and the in-game level-based adaptation in terms of self-reported valence increase. No effect was found between in-game level-based adaptation and the control group.

There are many potential reasons for the found pattern. The player-based adaptation proved successful in enhancing valence compared to the control group and the in-game level-based adaptation. The proposed system was tailored to the specifics of The Flow Experience and made use of the best predictors for valence increase. This adaptation therefore is unique for this game (i.e. taking the in-game context into account), but works through adapting to individual differences. The improvement was level-independent and exceeded the smallest effect size of interest for both comparisons. Again, there are practical considerations when defining such a target effect size: The recommendations by

Ferguson [277] for a medium-sized effect are based on the scientific need to make meaningful distinctions between groups and conditions. With a large number of players, even smaller effect sizes might become very relevant. The trade-off can be considered in regard to the expense: In this case, the implementation of a simple adaptation was accomplished via a single line of code which resulted in an immediate medium-sized effect. Depending on the complexity of adaptation systems and game data analysis, as well as the size of the player base, the smallest effect size of interest might be smaller or larger for a given project. The methodological process presented here shows potential in solving some of the existing problems in affective game design, especially in regard to conceptual uncertainties.

The level-based adaptation was chosen based on the results of Experiment 1 which indicated different valence-ratings in level 2 compared to the other levels. However, there was in fact no statistically significant difference between low and medium attack speed for level 2 in Experiment 1. It might be the case that a similar attack speed pattern was preferred in all levels all along, with a larger range of enjoyable attack speeds for level 2. The null results provide in any case practical guidelines: The in-game level-based adaptation represented a non-significant game balancing effort based on primary affective data. Balancing can be seen as an important consideration for any type of game, but strategies and mechanisms to balance have many implications [280]. It can be reasonably assumed that in the case of The Flow Experience, the game level was not a relevant factor in our aim to increase player valence.

While there are many player experience studies that mirror a similar approach to validate affective games (e.g. [194], [195], [262]), there are as of yet no standards and many studies reporting affective games are difficult to compare due to their fundamentally different approaches (e.g. [187], [192], [199], [205], [206]). Ewing et al. [195] proposed a similar 2-step experimental setup to ensure the validity of game material by explicitly testing the emotional evocative nature of Tetris. Similarly, Lui et al. [204] first created an emotional model following Pacman sessions and then used this information to successfully build an emotion-adaptive game. However, both of these studies integrated multiple physiological measures that are often expensive and hard to implement for game designers. Our results support this notion that a 2-step process can successfully create an enjoyable adaptive game experience by only using subjective affect data that is easy to assess and robust in its reflection of relevant aspects of player experience. It also adheres to the standards illustrated in Chapter 4, showcasing the utility of following such standards in a way that

would make affective game research more comparable and easy to follow.

This approach can solve conceptual problems, but it makes practical operations necessary, i.e. the determination of statistical relationships between emotions and game characteristics. Emotions can not fully be measured, only certain characteristics can be measured that have an unknown mapping to emotions [32]. While certain methodological approaches are well researched (such as the use of physiological arousal measured through heart rate or skin conductance and its connection to fear [252]), these are still being influenced by factors such as context and individual differences. Furthermore, some emotions that are interesting from a game design perspective (such as nostalgia, shame, or melancholy) have no known associations with any objectively measured instrument, but should still be used for research and design purposes. To overcome these problems, games can test emotional relationships themselves and identify influences of certain contexts or individual differences, thereby allowing for empirically based design for any emotion.

5.6 Conclusion

The aim of this chapter was to provide a real and practical example of the applied emotion design framework, to on the one hand evaluate it as a theoretical and methodological tool and on the other hand add to the body of work researching affect-adaptive video games, while adhering to rigorous methodological standards. A lot of care went into the study design to mitigate risks identified by the analyzed studies in Chapter 3 and offer a clear and hands-on example to integrate the robust methods of the emotion design framework (Chapter 4) into the development and research cycle of a video game.

Overall, the two-step evaluation provided empirical evidence suggesting the practicality of this approach. The resulting adaptation mechanisms not only fit well with the body of existing research (specifically concerning dynamic difficulty adjustments) but were successful in reaching a target emotion goal. In comparison to other work in the field, this study reaffirms that applied standards could provide more unified, comparable, and replicable studies to possibly in the future understand more about the effect of affect-adaptive video games. As it stands, this chapter adds to the notion that adaptive emotional experiences are a promising direction for game designers and researchers, not only to create new and bespoke experiences but also potentially to research fundamental theories of emotion.

It has to be acknowledged however that this is one piece of evidence, applying the emotion design framework to a very focused and simple example. Ideally, the framework should

be helpful not only to building affect-adaptation mechanisms but also to understanding the affective player-game interaction for any game and making use of the proposed principles beyond automatic adaptation systems.

Chapter 6

A Case for Emotional Output

"A child does not catch a gold fish in water at the first trial, however good his eyes may be, and however clear the water; knowledge and method are necessary to enable him to take what is actually before his eyes and under his hand."

— Harriet Martineau, *How to Observe Morals and Manners* [281]

Until this point, effort was put into clearing up theoretical misconceptions and potentially flawed practices in designing affect-adaptive games, as they are typically understood as "affective games". However, as described in Chapter 4, affective games can also be understood as any game that makes explicit assumptions about the affective relationship between a player and a game. Indeed, looking at games from such a perspective, it becomes clear that every game involves some kind of emotional relationship with its players. This also means that the concepts and methods that have been the subject of this thesis can be applied to all kinds of research concerning video games and their affective nature.

For example, one interesting use case is the analysis of human behaviour given an expectation of emotional experiences when playing a certain type of game. There is no doubt that emotions play an important role in human behaviour and while theoretical perspectives might argue about the precise nature of this role, it is generally agreed upon that emotions include behavioural components [28], [29]. Not only do emotions themselves lead to certain behavioural patterns, but people also make choices and show behaviours based on their expectation of the emotion they will experience [36]. This is of course no surprise - we watch sad movies because they are sad, not despite the fact [282]. This is no

different from our choice of video games: If we want to experience relaxation, we might play games marketed as "relaxing" or "cosy" that include calming music, graphics, and gameplay tasks [283]. Based on this train of thought, it is very easy to imagine that many people would make the conscious choice to seek out games associated with relaxation when in a state of stress. Or - to put it into the words of the Emotion Design framework - games that feature relaxing emotional output may be more sought after when people experience stress.

Such an assumption is based on quite an extensive body of research. People are very willing to play video games as a medium to relieve stress [284] or even argue stress relief as a main motivator to play games [285]. With a whole genre of games marketed as "relaxation games" or "cosy games" [286], there might also be no doubt about the specific nature of stress-relieving emotional output. However, as repeatedly shown in this thesis up to this point, making assumptions about affective relationships between players and games, without taking context and individual differences into consideration, may pose risks. According to the Emotion Design framework, we cannot assume universal associations between any type of emotional output and subjective experience. The main aim of the "testing phase" described in Chapter 4 and Chapter 5 was to establish the given statistical relationships first and based on the results inform design decisions. In order to not make inaccurate conclusions, it is necessary to assume that affective interactions between a game and a player are not universal and specific statistical associations are potentially unknown, as shown in Fig 6.1.

As a consequence, the previously made simple and obvious hypothesis becomes an interesting showcase of these core principles of the Emotion Design framework. According to the framework, we cannot assume that a dark and moody virtual room will inflict universal fear and in the same light, we cannot assume that cosy games are inherently associated with the actual experience of relaxation. Not only is the previously made hypothesis a question worth testing, but the process of testing it may uncover additional insights about the nature of the emotional player-game interaction and by extension the nature of human experience. The following study therefore represents an exemplary case of applying the Emotion Design framework to higher order questions, unrelated to the actual affect-adaptation in video games, but rather targeted at answering more general questions about human behaviour and emotions when interacting with video games of a specific nature. To achieve this, the following section will describe a study examining the

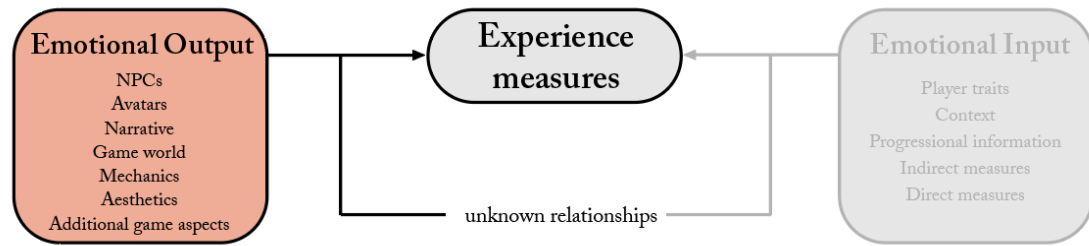


Figure 6.1: Illustration of the emotional output testing phase of the Emotion Design Framework. The core assumption of the framework is that the statistical relationship between game material and experience is dependent on context and individual differences and needs empirical testing. Assumptions about this relationship might lead to misguided conclusions.

demand for relaxing gaming experiences during the COVID-19 pandemic as a particularly stressful time, using relaxation-related emotional output descriptors in the form of Steam tags.

6.1 Background

In 2020, the COVID-19 pandemic had an infection rate of over 60 million cases [287], making it a global health crisis with many implications for mental health and psychological wellbeing. These are seen as a result not only of a direct infection with the virus, but rather overall social and personal difficulties arising from the crisis, such as financial losses, uncertainty about the future, and public health measures like social distancing and contact restrictions. Common mental health problems following the COVID-19 pandemic are often of an affective nature and include symptoms of anxiety, depression, and stress [288]–[290].

For example, Salari et al. conducted a meta-analysis involving 9,000 participants across multiple countries and found a stress prevalence of nearly 30% in the general population. Xiong et al. [290] similarly found many affective problems across America, Europe, and Asia connected to the pandemic, including increased rates of post-traumatic stress disorder (PTSD) and stress.

The adverse effects of stress on mental health have long been known, and there are many examples in the literature that describe the consequences of stress-inducing life events on

health and emotional well-being. Stress have been associated with pathological symptoms of depression [291], [292], anxiety disorders [293], [294], and even schizophrenia [295]. Considered from a different perspective, stress has also been shown to impact psychological well-being negatively, reducing life satisfaction [296], and happiness [297].

These developments are based both on biological [298]–[300] and psychological [301], [302] processes and represent a key topic of interest for the involvement of emotions in mental health.

Because of its strong associations with mental health, researching stress within the emotion-player relationship in video games might uncover potential benefits of further developing and researching affective games. Games already show strong potential in improving mental health through their effects on stress: For example, Russoniello et al. showed that casual video games (such as pinball) can reduce physical stress responses [303]; while Holmes et al. [304], [305] and Iyadurai et al. [306] provided empirical evidence for the potential of Tetris to reduce flashbacks in PTSD patients. Furthermore, many studies have been conducted that report general stress relief following the use of video games, with implications for improved mental health measures [25], [26].

Not only has it been shown that video games can relieve stress, it has also been repeatedly reported that people actively choose to engage with video games to recover from stressful situations [283], [284], even citing stress relief as a main motivation to play in the case of middle-aged adults [285]. This effect seems to be even more accentuated than the stress-relieving effect of other leisure activities, especially of passive media consumption [307]. Games that have "relaxing" effects (i.e. games with emotional output assumed to be connected to relaxation) are often marketed as such and consequently sought after. On the popular Steam platform for example, "relaxing" is one of the most popular tags, yielding over 2,800 game results in the Steam Store. These tags both represent attributes that are used by the developers to market a game and game characteristics identified by the player base, making them a potentially valuable descriptive attribute together with game genres in modern video game distribution platforms.

With the disruptions of everyday life activities during the COVID-19 pandemic and the pandemic itself constituting a major source of stress [289], [290], it seems quite believable that affect regulating aspects of video games become even more pronounced during lockdown. In fact, there are many reports indicating reduced physical activity and increased screen time in university students in the US [308] and Italy [309], adults in the US [310]

and India [311], as well as the general population in China [312], [313].

To summarize, it does not seem like a surprising assumption that people did in fact play more games (and especially games connected to relaxation) during the COVID-19 pandemic and therefore inherently facilitate the reduction of stress - at least in theory. But as previously discussed, this theory only holds true if we make the assumption of universal stress-relieving effects, given a certain type of emotional output. In terms of aesthetics, much effort has been made to systematically describe "cosiness" in games as an inherent attribute and how this attribute can explain psychological, physiological, and even societal developments [286]. To test if we can make such an assumption, given the context of the COVID-19 pandemic, the following study investigates the demand for relaxing video games during the COVID-19 pandemic by analysing Steam player data. To put it into a clear hypothesis: The peak number of daily players averaged over the time period March to November 2020 increased compared to the same time period in 2019 and significantly more so for "relaxing games" than any other type of game.

6.2 Materials and Methods

6.2.1 Dataset and Preprocessing

The study hypothesis and the analysis plan were preregistered on OSF (details can be found in [314]).

To operationalise a given game's demand, average daily player peaks (ADPPs) have been used as a measure of popularity for the time period of a day. The data set used includes excessive data from Steam, the largest online game distribution service for PC games with over 90 million monthly active users and 20 billion hours of gameplay in the observed time span [315], [316]. The data was obtained from SteamDB, an independent information service, accumulating data directly from the Steam API, including data about game followers, reviews, playtime, and concurrent players for every day and every game available on Steam.

The harvested data set initially contained records from 4,648 of the most played games on Steam. Information was collected about the name, developer, release date, and ADPP for each month since the release of each game. ADPPs were calculated for the time period March to November 2019 (pre-COVID-19 period) and for March to November 2020 (COVID-19 period). Because the COVID-19 pandemic developed differently over the world

and the study was conducted in December 2020, the pandemic period was defined as March to November 2020, based on reports indicating stress-related effects at the beginning of March [288]–[290] and the public considerations of stress-related concerns made by the World Health Organization [317].

Games that included missing values in ADPPs for any of the months in the observed time periods were eliminated beforehand, resulting in a sample of 4,147 games. Furthermore, games were only included in the next step if they were released before February 15, 2019, to ensure that each game had sufficient time to be available on the store to not skew ADPPs towards newly released games. This resulted in a reduction of the sample to 2,929 games. Finally, games with an average player peak of less than 3 players per day for the 2019 time period were excluded, resulting in 2,379 remaining games.

The game sample was then coded according to pre-defined coding criteria on a more detailed level. All games were coded in regards to their application type, as only games were to be included and no demos or other utility software. The games were assigned to the test group rather than the control group if they contained the steam tag "relaxing" and an analysis of the trailer, game description, and other tags did not lead to the conclusion that the game is not primarily relaxing. To fulfil the second condition, a game must not have contained any of the following: (a) Conflicting mood tags (such as "funny" or "emotional") with a higher priority than relaxing to ensure the primary emotional target experience was seen as relaxation and not sadness, melancholy, excitement, or humour. (b) Action-heavy and intense gameplay that shows clear signs of stress-inducing design characteristics (such as war-themed first-person shooters) as a primary experience characterisation of the game. (c) A main focus is on horror, sexual or mature content, or other characteristics that conflict with the relaxation aspects of the game. This process was conducted by both main researchers (see [3] for more details) and regularly checked for inter-coder agreement with further discussion on uncertain cases to prioritize games with a "cosy aesthetic" [286]. These efforts were made to compile a list of games that are primarily considered relaxing through their primary game loop and audiovisual presentation and not a list of games that are considered relaxing, even though the primary target experience is different.

Following this procedure, 143 games were identified for the test group (i.e. relaxing games) and 2,124 were assigned to the control group. Extreme outliers with more than 7 standard deviations from the mean in terms of ADPPs were further removed, leading to 2,216 games in total (138 relaxing games).

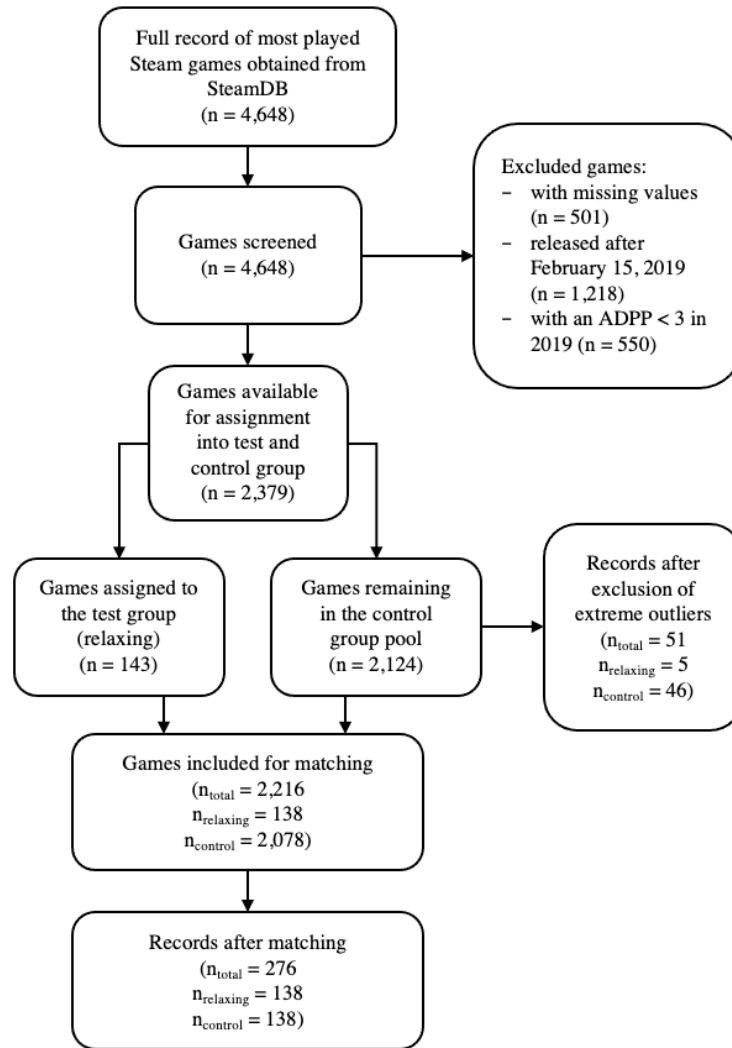


Figure 6.2: Flow diagram of the number of games screened and included in the study.

Finally, matching procedures were applied to enable further statistical analysis. Because the pool of relaxing games and control games were naturally different, a matching procedure was conducted to reduce covariate imbalance and therefore confounding factors. Pre-processes matched data is reported to produce more robust inferences tied to fewer assumptions, adjusting the samples based on a shared covariate [318]. The chosen covariate was the ADPP value in the pre-COVID-19 period (March to November 2019) as a shared starting point to observe changes for the COVID-19 period. Matching was achieved using a greedy nearest neighbour algorithm. In this method, each test unit is assigned one control unit based on the smallest distance in propensity scores between both units. Greedy matching was chosen as it takes the closest match for each value in the test

group, which has been shown to perform very efficiently with datasets containing a large pool of control units to choose from [319], [320]. The matching procedure led to a final dataset of 138 games in each control group and test group. A detailed overview of the selection process can be seen in Fig 6.2. A preliminary power analysis uncovered the need of 210 participants to reveal a medium-sized effect with a power of $1 - \beta = 0.95$. Because non-experimental data needs generally more conservative analyses, the final sample size of 276 was deemed appropriate to reveal a medium-sized effect.

6.2.2 Analysis

Data analysis was performed with the statistical computing software R. Both parametric and non-parametric analyses of variance were considered for the current dataset. Due to the violation of assumptions of parametric tests (e.g., ANCOVA, or change value ANOVA), non-parametric alternatives such as Quade's test and Wilcoxon signed-rank test were used.

6.2.3 Ethics Statement

Written consent was granted after reviewing the methods of our study by the Physical Sciences Ethics Committee of the University of York in a fast-track procedure. The full statement states: The researchers have taken all reasonable steps to ensure ethical practice in this study and I can identify no significant ethical implications requiring a full ethics application submission to the Physical Sciences Ethics Committee. I have checked and approved all supporting documents required for this application. I understand that completion of this form indicates that from an ethical point of view, I am willing to share responsibility for the work being conducted.

6.3 Results

138 games in the relaxing group were matched by the covariate (average daily players for 2019) with 2,077 games in the control group using the MatchIt package [321] in R. The final sample consisted of 276 games with 138 games in each group. Descriptive data regarding release years and game genres can be found in Table 6.1 and Table 6.3. Mean propensity scores for relaxing ($M = 0.06$) and control ($M = 0.06$) groups showed a standardized average mean difference of 0 and a maximal distance of 0.01. Fig 6.4 shows a visualization of the propensity scores for the matched and unmatched units. The mean

covariate value was $M = 226.19$ for the relaxing group, and $M = 226.19$ for the matched control group. A complete overview of descriptive statistics can be viewed in Table 6.4. Overall, nearest neighbour matching [322] resulted in a well-balanced dataset in regards to covariate similarity (see Fig 6.3 for the empirical quantile-quantile plot). However, due to the nature of the data acquisition, the distribution of the covariate was skewed for both groups.

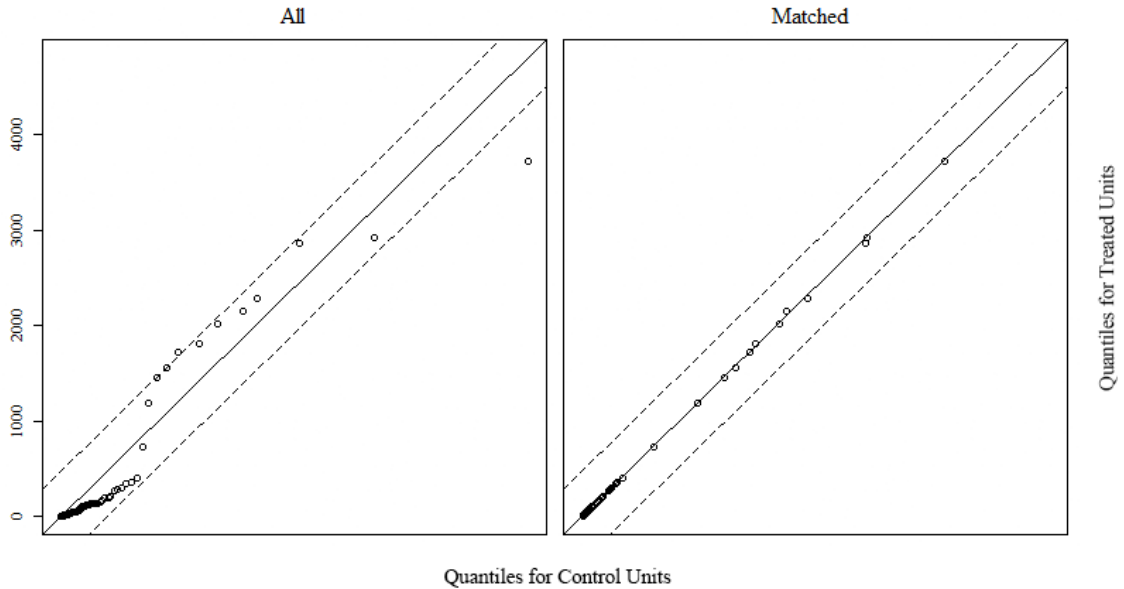


Figure 6.3: Empirical quantile-quantile plot of the covariate. Depicted are the quantiles of the covariate (ADPPs 2019) for the relaxing and control group before and after matching. Small distances between sample points and the diagonal indicate close similarity.

Group	2008	2009	2010	2011	2012	2013
Relaxing	1	0	2	1	2	9
Control	1	1	5	8	1	8
Total	2	1	7	9	3	17

Table 6.1: Release years of games in the relaxing group and control group (2008-2013).

Group	2014	2015	2016	2017	2018	2019
Relaxing	14	18	19	29	36	7
Control	15	14	17	36	30	2
Total	29	32	36	65	66	9

Table 6.2: Release years of games in the relaxing group and control group (2013-2019).

To test whether relaxing games and control games showed a similar expected ADPP

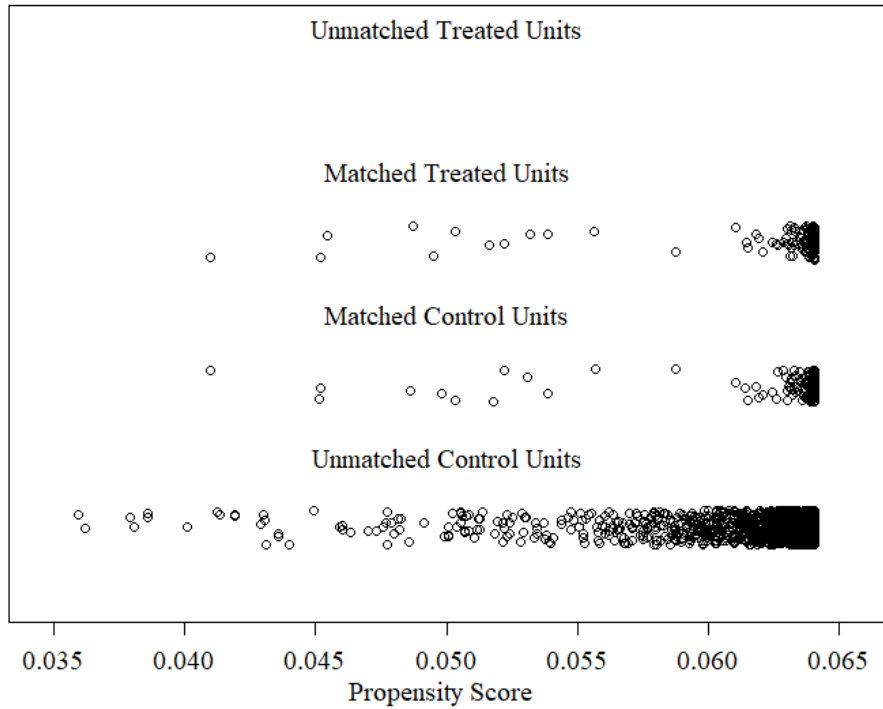


Figure 6.4: Propensity scores of matched and unmatched samples. Relaxing games are the treated units and control games are the control units.

Genre	<i>Relaxing</i>	<i>Control</i>
<i>Action</i>	1	91
<i>Adventure</i>	31	11
<i>Casual</i>	35	9
<i>Horror</i>	0	5
<i>Puzzle</i>	10	0
<i>Racing</i>	1	1
<i>RolePlayingGame</i>	5	8
<i>Simulation</i>	46	3
<i>Sports</i>	2	3
<i>Strategy</i>	7	7
Total	138	138

Table 6.3: Genres of games for the relaxing group and the matched control group.

progression based on time preceding the pandemic (2015-2019), a preliminary linear regression analysis was conducted, predicting ADPP by time and group. There was no significant main effect for time ($t[752] = 1.67, p = .10$) and no interaction effect between time and group ($t[752] = 0.55, p = .58$). A graph of the progression of ADPPs over time by group can be viewed in Appendix A.1.

Preliminary tests of ANCOVA assumptions using scatterplots showed a sufficient lin-

		2019		2020	
<i>Groups</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Relaxing</i>	138	226.19	612.16	246.54	652.76
<i>Control</i>	138	226.19	657.38	240.43	657.39
Total	276	226.19	611.26	243.49	653.89

Table 6.4: Sample size (N), mean (M) and standard deviation (SD) for the ADPPs in 2019 and 2020 by groups.

ear relationship between the independent variable (average daily players for 2020) and the covariate (see Fig 6.5). Regression slopes for both groups were parallel, pointing towards homogeneity, which was tested using a two-way ANOVA for the interaction between covariate and group ($F[1, 272] = 0.05, p = .82$).

Results showed no indication of an interaction, so homogeneity of regression slopes was assumed. Just like for the covariate, distributions of the dependent variable and its residuals were highly skewed, which violates not only basic assumptions for ANCOVA but for a potential change value ANOVA, which is commonly used as an alternative to present average treatment effects on treated (ATTs), i.e. in this case the gain of average daily players of the relaxing games group specifically between 2019 and 2020.

It was therefore decided to explore the overall time effect and the ATT using more robust non-parametric methods that are not distribution dependent. First, the overall time effect (the increase in ADPPs for both groups combined between 2019 and 2020) was calculated using a Wilcoxon signed-rank test [323]. The test showed a significant difference between 2019 and 2020 ($Z = 4.70, p < .01$). So, between March 2020 and October 2020, a significantly higher daily average peak per game was observed than for the same period in 2019. Mean ADPPs increased from $M = 226.19$ in 2019 to $M = 243.49$ in 2020 (see Table 6.4 for more information). Effect size was calculated to quantify the interpretation of the effect after Rosenthal [324]. Using Cohen’s criteria [325], the calculated effect size ($r = 0.20$) can be considered a medium-sized effect.

To test if this increase was even larger for the relaxing games compared to the control games, the ATT was calculated using Quades’s rank analysis of covariance [326]. This method utilises the residuals of the regression of ranked dependent variables and ranked covariates, rather than the non-normal distributed variables themselves. Results showed no significant group effect ($F[1, 67] = 0.05, p = .88$), meaning the ATT was not significantly different from the average treatment effects on the control group. In other words: No significant difference of average daily player increase was observed between the relaxing

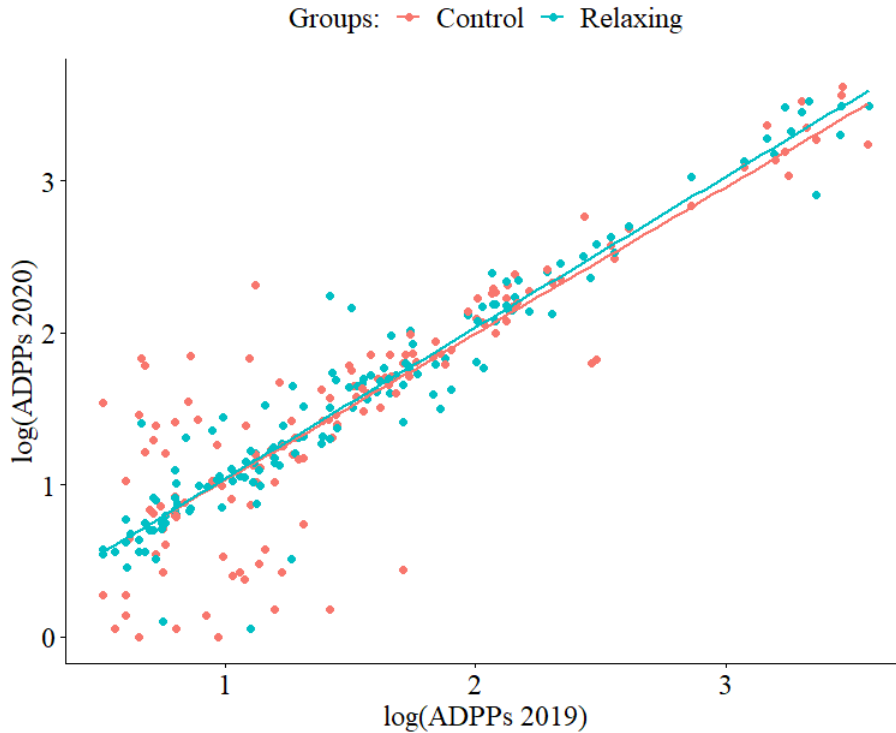


Figure 6.5: Scatterplot of the relation between ADPPs in 2019 and ADPPs in 2020 by groups. Scales have been log-transformed for clearer data presentation.

game group and the control group. A boxplot visualising the group differences can be seen in Fig 6.6.

6.4 Discussion

To examine the increase in demand for video games associated with relaxation during the COVID-19 pandemic, relaxing games and matched non-relaxing games have been compared in regards to their average daily player peaks for the periods of March to October in 2019 and 2020 respectively. Results revealed a medium-sized increase of ADPPs over both groups for the COVID-19 period compared to the pre-COVID-19 period, but no significant differences between both groups regarding this increase.

While the experiment uncovered an overall increase in ADPPs, there was no evidence to support the special role of games associated with the "relaxation" tag during the pandemic. There are various implications for these findings regarding affective classifications of game material (i.e. emotional output according to the Emotion Design framework) and its

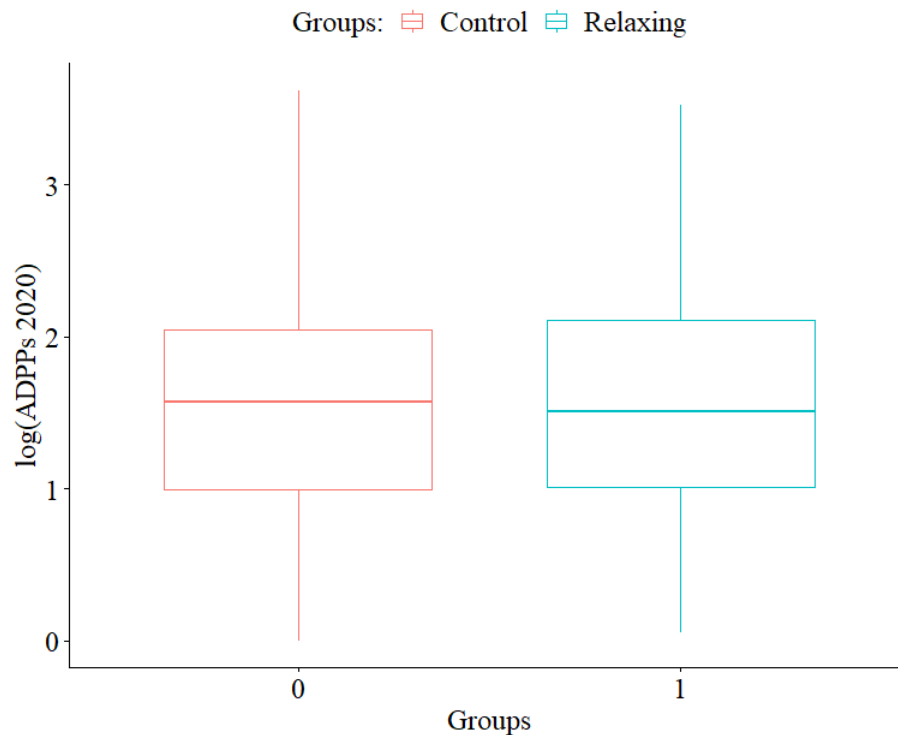


Figure 6.6: Boxplots of the differences in ADPPs in 2020 by groups. Scales have been log-transformed for clearer data presentation.

implications for observed behaviour within the affective player-game interaction.

6.4.1 The Role of Relaxing Games in Increased Daily Player Peaks

The COVID-19 pandemic has had vast influences on people’s everyday behaviour and the presented increase of ADPPs in the current study mirrors the current findings in the literature, including the findings of the increase of screen time all over the world (e.g. [308]–[312]). During the months of lockdown, availability for outdoor activities was very limited and digital alternatives, including video games, have experienced a rise in popularity. In fact, there is no reason to doubt an overall increase in digital activities, but there is disagreement regarding the reasons and consequences of these behaviours, specifically for video games.

Just as there is not much doubt about the behavioural changes that resulted from the pandemic, the literature provides a huge amount of convincing evidence regarding the changes relating to mental health, specifically psychological stress [288]–[290]. Multiple studies associated COVID-19-related increased screen time and decreased physical activity

with mental health problems (e.g. [327], [328]). The pandemic itself and the consequences of the lockdowns on peoples' lifestyles have been identified as risk factors for psychological problems. However, more and more recent studies provide arguments that link the role of video games in specific not to a mental health danger, but to mental health improvement and protections against such risk factors (e.g. [25], [303], [304]). It further seems that people are more and more drawn to video games as a way to actively recuperate from stressful events [283], [285].

While the present findings do not hold any information about why people play more during the pandemic, the increased numbers of daily players across all games do not contradict this particular view. It is however important to note that this overall increase might not be related to COVID-19 but rather represent a naturally expected increase of ADPPs independent from the pandemic, but possibly related to factors such as reduced cost or increased accessibility of technology. As Appendix A.1 shows, ADPP peaks for most games seem to lie in 2018 with a big decrease in ADPPs for all games between 2018 and 2019. It is very much possible that the pandemic affected this trajectory from 2019 onwards, inverting this trend. The overall effect may however not be as significant as it seems considering the the overall variation in ADPPs over time and the unexplained decrease of ADPPs after 2018, which might represent multiple reasons, such as a natural decrease in interest after a few years post-release. In any case, the data does not allow for conclusive attribution of observed trends. Not only is it questionable to assume the pandemic had a large effect on the player increase, but the presented data shows no support for the assumption that people would play even more games associated with relaxation to cope with the stress caused by the pandemic.

Looking at player demand, there seems to be no significant difference between "relaxing" or "cosy" games and any other type of game. The natural conclusion to make is that our presented collection of relaxing games does not represent an objective list of relaxing games - or to put in other words - games commonly associated with relaxation (by both developers and players) may not have a universal association with the actual subjective emotional response of relaxation. This study was carefully conducted with objective criteria and replicability in mind and can therefore stand as an indication that we are in need of a broader discussion about what a "relaxing game" is in a scientific context. As can be seen in Table 6.3, the test game group included a specific type of game, representing genres such as simulation, adventure, and casual games - in recent years often described

as "cosy" [286]. These are games that are categorized by the players as relaxing but are not necessarily correlated with the actual experience of relaxation. The current results therefore represent the core principles of the Emotion Design framework (see Fig 6.1): The actual experience of relaxation given a certain type of emotional output is highly dependent on individual differences and context. Playing "relaxing" games means different things for different people in different circumstances. Furthermore, this finding might also represent a disconnect between game labels and the actual affective phenomena these labels are based on - speaking in a broader sense of emotional experience in general. This disconnect is very naturally bound to the conceptual uncertainties regarding emotions (and most affective phenomena for that matter) and the differences in viewing and describing emotions between psychologists, designers, computer scientists, game publishers, players, and other subgroups of the general public.

6.4.2 Means of Relaxation

It would be a fallacy to assume that the presented findings provide evidence against the stress-relieving effects of video games. But rather instead of seeing relaxation as a universal game characteristic, the actual experience of relaxation emerges from an interaction between a game and a player. What is labelled as "relaxing" by a general population might not hold much weight, seeing that emotions are inherently subjective. A wide variety of individual differences regarding physiological stress responses have been reported [329] and emotion regulation technique effectiveness is also known to be highly dependent on individual differences [330], [331]. Game elements that consistently fall under the description of "relaxing" might therefore not mirror what individual people find relaxing and seek out.

One source of confusion could be the amount of mostly violent action games in the control group (see Table 6.3), which could be thought to cause stress, rather than reduce it (e.g. [332], [333]). One explanation for why this might not be the case is the complexity of games itself. Given all games have so many factors and elements connected to them, it may hardly be possible to relate one experience to one game. For example, many very violent games integrate some form of social interaction, which has been identified as an important factor moderating positive effects of video games on mental health [334], [335]. In fact, social interactions have consistently been shown to be an important buffer against the negative impacts of stress (e.g. [336]), which also relates to social components within otherwise violent video games. Given the number of potential moderators on the effect

between video games and emotional response, it may not be surprising that some explicitly violent games seem to show similar trends to explicitly relaxing games.

Still, on a more fundamental level, many researchers argue that specific game elements do not completely account for the observed experience in a universal way. More recent research has been conducted that demonstrates the importance of individual differences in response to violent games (e.g. [335], [337]) and even sheds light on important elements with a positive relation to mental health. For example, Collins and Cox [334] found that action games and shooters have the largest effect on after-work recovery than any other genre. What therefore constitutes as an emotional output connected to the actual experience of relaxation might not follow "relaxing" or "cosy" aesthetics. For some people, games connected to actual relaxation can consist of quite contrary characteristics. Maroney et al. [338] argue that all types of games provide means to reduce negative states, depending on individual stress-coping mechanisms, and that social interactions could be understood as mediators for these individual effects. Again, the evidence points towards the importance of individual and contextual differences when experiencing emotions and therefore - in line with the Emotion Design framework - the importance of explicit testing of assumptions regarding affective associations between game material (emotional output) and actual emotional response. As previously discussed, affect proves to be complex, ambiguous, and non-universal - as it has been described by modern psychologists for years [28], [36], [37]. This study can therefore be seen as a supporting argument for the need for the Emotion Design framework as a way to accurately illustrate affective game-player interactions and a guide to robustly provide affective experiences within video games.

In many current studies, a high emphasis is put on how specific games with relaxing elements could be associated with well-being. Specifically games like *Animal Crossing: New Horizons* have been used as an example of the possible effects of video games during the pandemic [339], [340]. While this study does not necessarily invalidate this approach, it might be worthwhile to expand the research to include all kinds of games, game elements, styles, and genres and look at their interactions with individuals to elicitate actual emotions, as described in Chapter 4.2.

6.4.2.1 Limitations and Future Directions

Some limitations have to be considered that emerged as a consequence of the study design. As no randomization was possible and a quasi-experimental design was chosen, it might be

the case that a naturally lower statistical power could have influenced the reported results. Measures like the reported effect size of the overall ADPP increase must be interpreted with caution, as these could behave differently (and often more conservatively) than the respective measures for parametric tests. The limited sample size also limits the probability of revealing very small effect sizes that still could be relevant for data containing video game players all over the world. To have a more precise picture of the discussed effect, a large-scale experimental study would be appropriate, although the informational gain might not be worth the effort. Another characteristic of this study is the restriction to Steam and a certain player base. Steam is only a part of the whole video game industry and does not reflect players on consoles or mobile devices. Because relaxing games were limited in ADPPs and controls were matched, most of the most popular games with hundreds of thousands of daily players had to be excluded. As a result, this study provides insights about a specific type of game for a specific platform. It does not reflect differences in types of players, which could also be an interesting topic for further examination. Methodological limitations are however a consequence of this data-driven approach, which is essential to explicitly test affective associations. Despite limitations, it stands as an example of the knowledge gained when conducting empirical studies - even for seemingly obvious theories. It would be misguided to conclude that the demand for relaxing games is currently equal to the demand for non-relaxing games. Instead, the results should be interpreted in regards to their informational value when it comes to the gaps that are currently present when talking about affect-eliciting game elements and an example of applying the Emotion Design model to an empirical study to further our knowledge of human affect.

6.5 Conclusion

The Emotion Design framework can be seen as a guide to developing affect-adaptive video games, but also provides theoretical lenses to understand affective phenomena in user-centered software. This chapter provided an empirical experiment investigating a real-world phenomenon (i.e. player behaviour during the COVID-19 pandemic) through the concept of emotional output. Understanding emotional output as game stimuli with the potential to elicit emotional reactions for a certain type of player in a certain context can help understand behavioural patterns in large parts of society. It is not enough to assume a game or game characteristic is relaxing - the label itself only derives meaning if the actual affect is experienced by the players.

Chapter 7

A Case for Emotional Input

"Practically speaking, then, there is no "gold standard" measure of emotional responding."

— Iris Mauss & Michael Robinson, [32]

Chapter 6 discussed the concept of emotional output and how this label in the context of the Emotion Design framework can help explain certain observations regarding the affective game-player interaction. This is true for various kinds of games that are not traditionally seen as "affective games" [46] as they do not necessarily concern themselves with adaptation on primary emotional data. Still, the theoretical and methodological barriers thoroughly discussed in Chapter 3 also apply to games that process user emotions. Emotional input as a concept relates to player information that could be used to infer affective states. Crucially, there is no one-to-one mapping between a certain measure and an emotional state [32] so similarly to emotional output, emotional input represents context-specific information that differs between individuals. The following chapter demonstrates the usefulness of such a conceptualisation in explaining emotional phenomena observed in real game scenarios.

As Fig 7.1 shows, game systems might have access to a variety of potential user information, such as user input, their in-game abilities and progressions, and even something like traits or psycho-physiological measures. Many studies have argued that especially biofeedback instruments are successful tools to model player emotion [40] and consequently provide sufficient information for a game system to adapt. When looking however closely at real video game adaptation, as shown in Chapter 3.3, we can observe that emotions are modelled and measured with a variety of different instruments for a variety of different games - often making specific assumptions about how certain outcomes relate to affective

states. For example, I found that many studies used similar physiological measures (such as HR and EDA) to assess vastly different emotional states relating to both dimensional and distinct underlying structures. There is a substantial body of research demonstrating the utility of such instruments to assess affective information [252]. In the context of the Emotion Design framework (see Chapter 4.1.3) it is important to view this information as dependent on context and individual differences. In other words, it becomes necessary to test the predictive power of a given instrument for a specific target emotional state with these possible influences in mind.

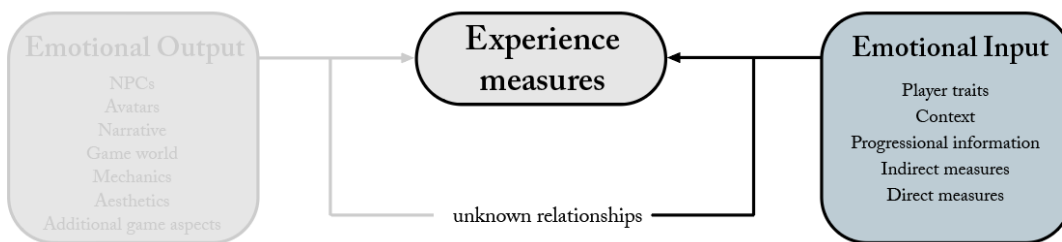


Figure 7.1: Illustration of the emotional input testing phase of the Emotion Design Framework. It is assumed that as with the emotional output, relationships between player input (gameplay, physiology, movement, etc.) and subjective affect are dependent on context and individual differences.

The Emotion Design framework guides the process of testing both the statistical relationship between emotional output and target emotion and emotional input and target emotion. The ultimate aim is a validated closed loop: The emotional output influences player emotions that we can measure within the game through emotional input, which in turn can be used to change the emotional output, and so on. Affective information used in this loop needs to be clear and tested, as was shown throughout this thesis. The Emotion Design framework can be a useful tool to achieve this.

For example, the affective valence of players through an immersive virtual reality game is, according to the framework, at least partly dependent on the player and their circumstances. In this chapter, this scenario is explicitly tested by modelling the affective journeys of players of two distinct groups: A group consisting of sleep-deprived players and a sleep-rested control group. Sleep deprivation has been related to differences in anxiety expression [341], while being a purely circumstantial and temporary characteristic of cer-

tain players in this experiment. Any studies assuming objective, fixed pathways between measurement instruments and observed emotions (see Chapter 3.3.2 for examples) might assume a difference between these groups in terms of affective measurements, but no difference between these groups in terms of the predictive power of certain instruments. But if not just the emotions of players change depending on their level of sleep but also the statistical relationship between measures and observed emotions, this assumption would be wrong and further demonstrate the need of a more robust conceptualization as provided by the Emotion Design framework.

This chapter uses a fully immersive virtual reality horror world based on the Underwood Project by McCall et al. [342]. The virtual world represents a dark interior structure full of rooms with anxiety-associated stimuli, connected through narrow pathways. Players assumed a first-person perspective and moved around the world, while multiple measures were used to assess their affective state, including physiological measures (heart rates and electrodermal activity) and position measures (position and rotation of hands and head). The VR horror world and data gathering was conducted as part of a previous study on sleep deprivation [343]. The present study made use of that data to perform a novel set of analyses to test contextual influences of predictive emotion models. The participant's affective journeys throughout the game was modelled based on self-reported valence as an outcome. The main hypothesis of this study was that the predictive model (i.e. the linear regression between measures and outcome) would be different between the sleep-deprived and sleep-rested groups.

7.1 Background

Sleep deprivation can be seen as an individual circumstance with a potential impact on affective reactions. For example, state anxiety, as a form of short-term affective apprehension and nervousness reactions to triggering events [344], has been shown to be influenced by sleep deprivation compared to full-night sleep rest [341], [345]. In a recent study by Sullivan [346] these findings were again expanded by showing that sleep deprivation led to both heightened subjective anxiety and physiological anxiety reactions (measured through HR and EDA) during a VR horror game experience involving ambiguous threats (i.e. threats of unclear nature [342]).

These findings showcase that sleep deprivation has a significant impact on how emotions - specifically anxiety - are expressed and experienced. This again reinforces the notion

that individual differences and context are important factors when it comes to explaining emotion elicitation and expressions in individuals. While this alone has important implications for affective video games, it is not yet clear if these factors also influence how affective reactions can be predicted. As shown throughout this work, there is no objective measure of emotion [32], but affective video games still make use of various instruments (including biofeedback instruments [40]) to model players' emotional journeys throughout games. In the context of the emotion design framework [2], this can be seen as emotional input. This concept was introduced to both label measurement instruments in a useful way to describe their role in the affective feedback loop when developing games and to underline the importance of validating these instruments given a target experience, specific game, and target audience.

Not all affective games validate these measures [1], potentially because of the previous lack of methodological standards and theoretical clarity. But if player characteristics - such as sleep deprivation - could influence not only emotional reactions but also the statistical relationship between predictors and outcome emotional reactions, the need for such validation efforts would become even more evident. To test this hypothesis, this study expands the study by Sullivan [346] by testing if a sleep deprivation vs. sleep rest condition would interact with affective predictors (such as HR, EDA, position) when predicting emotional subjective valence in the same VR horror game experience.

7.2 Materials and Methods

7.2.1 The Underwood Project

The Underwood Project is a virtual reality (VR) environment, originally designed and tested by McCall et al. [342] as a tool to elicit and research ambiguous fear. Created using the authors' modular Unity 3D kit, the VR horror world in the present study was created as a short First-Person VR horror game with the aim to move through a linear world filled with design elements based on ambiguous threat [346], including ambient audio [347] and audio events indicating the presence of unseen agents [348], visual stimuli associated with threat, such as blood, medical equipment, insects [349], [350] and restricted lighting, only providing glimpses of the world through the character's torchlight and lanterns.

The aim of the game was simply to navigate the in-game character to the end of a linear path made up of various corridors, rooms and two elevator sections. There were

two types of rooms: First, there were "neutral" rooms that were designed as plain offices, without the design elements used to elicit ambiguous fear [342]. The second type was anxiety-inducing rooms, often containing anxiety-related stimuli of a specific theme (e.g. hospital room, autopsy room, etc.). When players enter these rooms, the connecting doors close and players are trapped until they collect a lantern that appears somewhere in the room by simply touching it with their in-game character.

Players start in the upper level (neutral zone), which is brightly lit in one of the neutral office rooms and then enter an elevator that brings them to the dark and atmospheric basement level, where they encounter the first set of three consecutive anxiety-inducing rooms, followed by a set of three neutral rooms, and again followed by the second set of three anxiety-inducing rooms. Afterwards, the second elevator takes them back to the upper level and the game ends when they enter the last neutral office room. A complete description of the experience, including a video showcasing each section can be found in the study by Sullivan et al. [346]. One session takes approximately 10 minutes.

This version of the Underwood project was built using the modules provided by McCall et al. [342] and enhanced through various visual and auditory effects such as creepy sounds, insects, blood on the walls, moving shadows, self-closing doors, etc., as well as custom scripted interactions (e.g. the appearing lanterns that had to be collected). The world was built in Unity 2020.3.21f1 and scripted with C# with additional 3D models developed in 3ds Max 2017.

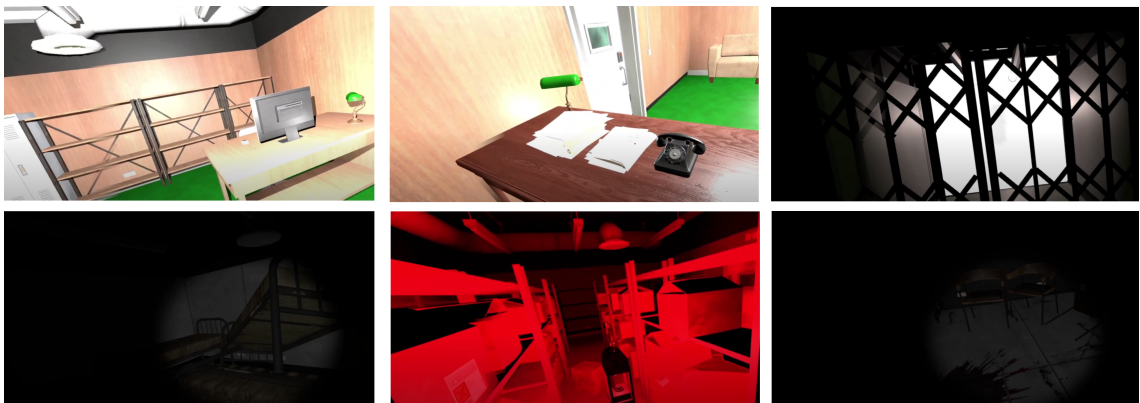


Figure 7.2: Screenshots of the Underwood VR experience. Top row left to right: Neutral starting room, neutral recovery room, lift to basement level; bottom row left to right: anxiety-inducing starting room, anxiety-inducing room with a lantern, late anxiety-inducing room.

7.2.2 Measures

This study made use of physiological equipment to measure players' heart rates (HR) and electrodermal activity (EDA) during the play session. HR was measured through a wireless Biopac BioNomadix electrocardiography (BN-RSPEC) amplifier with a three-lead set of pre-gelled, disposable Ag/AgCl foam electrodes (Biopac, EL503). The electrodes were positioned on the right clavicle, left mid-clavicle, and lower left rib cage. EDA was measured through a wireless Biopac BioNomadix amplifier (BN-PPGED) with a BioNomadix dual electrode lead, and ungelled, disposable Ag/AgCl foam electrodes (Biopac, EL507a). The electrodes were attached to the middle phalanges of the left middle and index fingers using an isotonic electrode paste (Biopac Gel 101a). The physiological signals were recorded using a Biopac MP160 acquisition system (2000Hz) and AcqKnowledge 5.0.

During the game, various player and in-game variables were continuously recorded. First, the position and rotation of two VR controllers (one per hand) and the headset were measured for every frame of the game running at 60 frames per second (FPS) as a measure of players' positions. For the hand controllers and headset, three values were calculated to represent the change of position compared to the original hand and head positions at the beginning of the game on three axes (x, y, z). Similarly, rotation was represented within three values, indicating the change of rotation of each controller and the headset compared to the original rotation in degrees on the x-, y-, and z-axis respectively. The Euclidean distance for each position and rotation variable from the origin point (0,0,0) has been calculated at each time interval. The Euclidean distance provides a scalar representation distance in 3D space, which is essential for quantifying movement characteristics.

The formula used for this calculation is as follows:

$$d = \sqrt{x^2 + y^2 + z^2}$$

with d representing the distance compared to the origin point. The position of the in-game character within the game was also tracked in three values representing the x-, y-, and z-axis and the Euclidean distance has been calculated in relation to the centre of the digital world. A complete list of predictors can be found in Table 7.1.

Subjective affective experience was measured using an implementation of the Affect Grid, originally proposed by Russel [276]. It includes an axis for valence and one for arousal as dimensions of subjective experience (or feeling [29]). For this study, only the valence axis was used as a measure of player experience [268].

Type	Predictor	Aggregation Function
Context	Group	None
Context	Time	None
Biofeedback	EDA	Mean
Biofeedback	HR	Mean
Position	Right Hand Position Distance	Sum
Position	Left Hand Position Distance	Sum
Position	Head Position Distance	Sum
Position	Right Hand Rotation Distance	Sum
Position	Left Hand Rotation Distance	Sum
Position	Head Rotation Distance	Sum
Ingame	Player Position Distance	Sum
Ingame	Player Rotation Distance	Sum
Individual Differences	Participant ID	None

Table 7.1: List of predictors by type and the corresponding aggregation function for the down-sampling step.

7.2.3 Procedure

The overall procedure spanned over three different sessions and a detailed overview can be seen in the study by Sullivan [346]. In the first session, participants provided their demographic details, filled out screening questionnaires and then entered the VR world with the physiological equipment attached to test if they would be comfortable with the experiment and if the physiological data would be usable for further analyses.

At least 24 hours later, session two started (around 8:30 pm). Participants were randomly assigned to either the sleep-deprived group or the control group (see section 7.2.4 for more details). The control group was tested overnight with polysomnography to ensure they had a sufficient amount of sleep. Polysomnography was achieved using the Embla N700 polysomnography (PSG) system. At approximately 11 pm, lights were turned off and at 7 am the next morning, participants were awoken again. The experimental group (sleep-deprived group) was not permitted to sleep during this time but was able to spend the time to their leisure with the addition of a questionnaire that involved the answering of general knowledge questions every 30 minutes until 6:30 am. Participants in this group also had to wear an actigraphy device on their wrist to ensure that they would not be sleeping during the night.

The final session started in the following morning (8:30 am). Participants were connected to the physiological measures and had a 5-minute resting period which was followed by a full playthrough of the VR horror game while their EDA and HR were measured. The game was played using a HTV Vive headset with an integrated Dual AMOLED 3.6-inch

screen per eye (1080 x 1200 pixels resolution), a refresh rate of 90 Hz, and a 110-degree field of view. The game's audio was presented through DOQUAS wireless headphones. After the playthrough, each participant viewed a screen recording of their play session and was instructed to continuously rate their subjective emotional experience using an implementation of an affect grid [276] within the open-source software DARMA [351]. This implementation used an XBOX wireless controller joystick, so players were able to continuously indicate their valence and arousal levels during the game.

7.2.4 Participants

Participants were recruited using the participant system of the University of York. In the screening session (session one), participants had to fill out a variety of questionnaires to gauge their fit for the study and were then excluded if they met one or more of the following exclusion criteria:

1. A history of neurological, psychiatric, attention, or sleeping disorders.
2. Scoring over the clinical cutoff for either anxiety as assessed by the Beck Anxiety Inventory (BAI; [352]) or depression as assessed by the Beck Depression Inventory (BDI; [353])
3. Extreme diurnal preference (score of >69 or <31) measured through the Morningness-Eveningness Questionnaire (MEQ; [354])
4. Poor sleep quality (score of >6) according to the Pittsburgh Sleep Quality Index (PSQI; [355])
5. a regular bedtime of 2:00 am or later

An additional inclusion criterion was a regular rise by 8:00 am in the morning. 85 participants completed the initial screening session, with 64 completing the main experimental sessions. Due to violating the experimental protocol, 9 participants were excluded from the analysis and one additional participant was excluded because they reported nausea during the VR section of the experiment. The final sample therefore consisted of 54 participants (33 females) with a mean age of 19.95 years ($SD = 2.17$). Participants received either £90 as a payment or course credits from the University credit system. Participants were randomly allocated to either the sleep-deprived or sleep-rested group.

7.2.5 Analysis

All measures were pre-processed as described by Sullivan [346], most notably down-sampling of the data to 500Hz and removal of artefacts > 2 seconds. Each predictor variable was further down-sampled to reach 1Hz based on similar approaches [115] by applying the corresponding aggregation function presented in Table 7.1. Specifically, biofeedback data was aggregated by calculating the mean value within 1 second, distance data was aggregated by calculating the sum (i.e. overall distance travelled) within 1 second and individual differences and context stayed constant for the time aggregation. This led to 31,762 observations altogether with 54 observations (one for each participant) per time point (1s). Correlation coefficients between each predictor and valence were calculated. Left-hand rotation and Player Rotation were excluded from further analyses as the recorded values were observed to be constant due to an error in the measurement setup.

To reduce the dimensions of the remaining predictors, a principal component analysis (PCA) was conducted. Factor loadings were used to reduce predictor dimensions into orthogonal principal components using a Varimax rotation, ensuring each new component captures unique aspects of the data. The cumulative explained variance was then plotted in a scree plot using the new principal components and used to determine which components were used in further analysis based on a cutoff of 80% explained variance [356], [357].

The scores on the remaining components were then further aggregated by calculating mean values per participant per room of the game world. This means that the time variable was first transformed to become a categorical room variable starting from the moment a room was entered and ending when the next room was entered (or the game ended). As described in Section 7.2.1, rooms were designed to be neutral or scary and were therefore deemed to be a meaningful representation of participants' emotional journey within the game world. Principal component scores were then aggregated to have one value per participant per room, providing a total of 487 observations.

Finally, the data was used to create a Mixed Model with a random intercept for each participant to account for within-participant variation and a random intercept for each room to account for between-participant variation within rooms. To test the potential influence of interaction terms with the group condition (sleep-deprived vs. sleep-rested), the mixed model additionally included interaction terms of each predictor (PCA scores) with the group condition. This was done to analyse the influence of the individuals' context or circumstance when it comes to predicting valence based on biofeedback and position

measures within a horror game world. All calculations were done in Python with the scikit-learn toolkit [358].

7.2.6 Ethics Statement

This study was approved by the Department of Psychology Research Ethics Committee at the University of York and all participants provided their written consent to participate.

7.3 Results

Results of the correlation analysis between predictors and valence can be seen in Table 7.2.

Predictor	<i>Pearson's r</i>	<i>p</i>
Group	-0.03	< 0.01
Time	-0.15	< 0.01
EDA	-0.12	< 0.01
HR	-0.04	< 0.01
rightHandPosition	0.33	< 0.01
leftHandPosition	-0.13	< 0.01
headPosition	0.33	< 0.01
rightHandRotation	-0.01	0.05
headRotation	0.08	< 0.01
playerPosition	0.33	< 0.01
participantID	-0.04	< 0.01

Table 7.2: Correlations between each predictor and valence with corresponding p-value.

Results of the PCA can be seen in Fig 7.3. Factor loadings reveal that position data of right-hand position, left-hand position, head position, and player position load onto the same component (PC1). Head rotation and right-hand rotation load negatively on PC2. Biofeedback data (EDA and HR) load on PC3 and PC4, with PC3 representing the shared direction of EDA and HR and PC4 representing diverging directions. The scree plot shown in Fig 7.4 shows the individual explained variance of each component. 80% explained variance was reached after including 4 of the 8 components (cumulative explained variance ratio after 4 components was 0.83).

As shown in Fig 7.5. both participant groups showed similar emotional journeys throughout the experience. In the brightly lit neutral rooms, valence values were on the higher end of the scale (20 to 60), while in the anxiety-inducing basement rooms, valence levels were on the lower end of the scale (-20 to -60). This pattern was true for all neutral sections (beginning upper level, mid-section neutral rooms, and final upper level) and

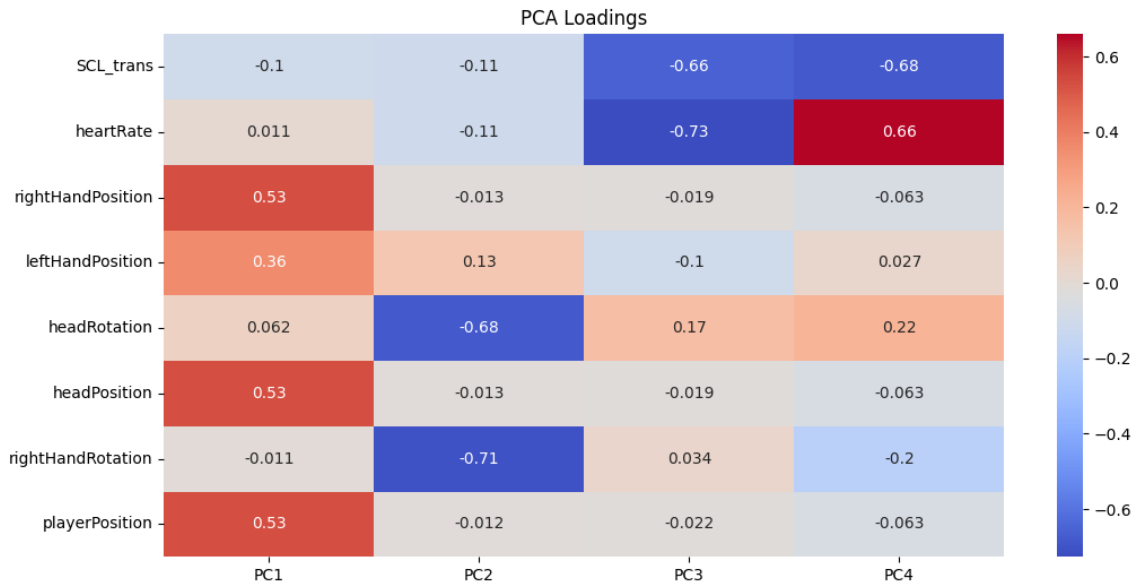


Figure 7.3: Illustration of the PCA results. The X-axis shows the new components, the Y-axis shows the original predictors.

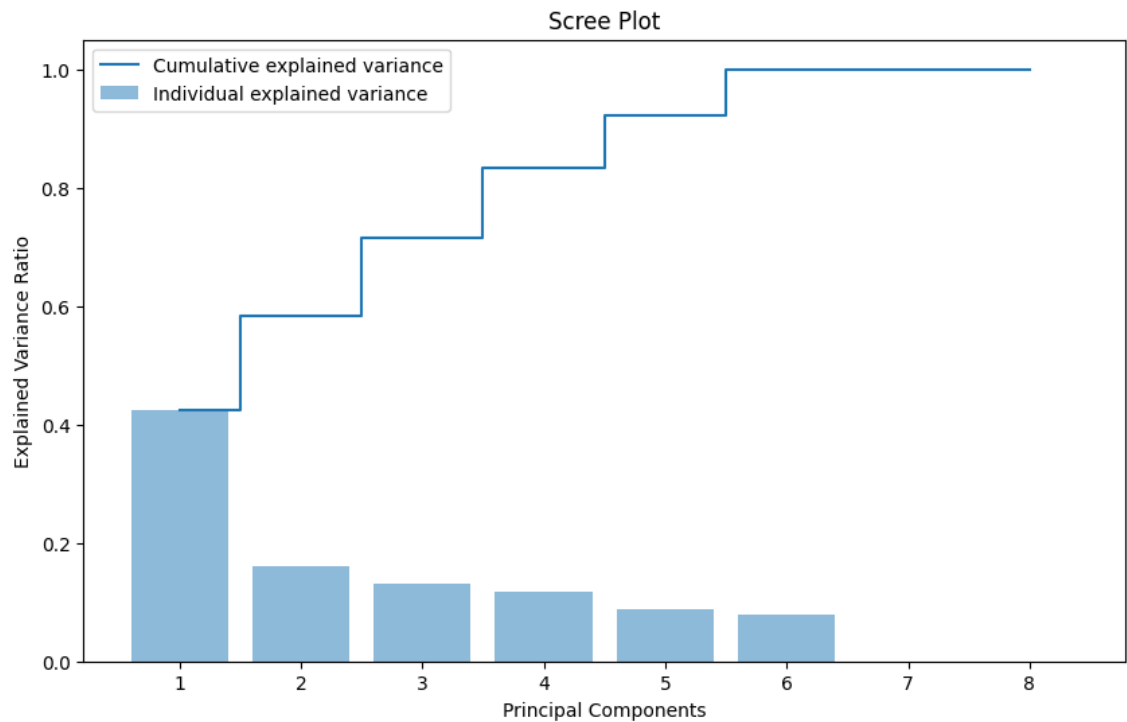


Figure 7.4: Scree plot showing the explained variance ratio of the individual principal components.

anxiety rooms.

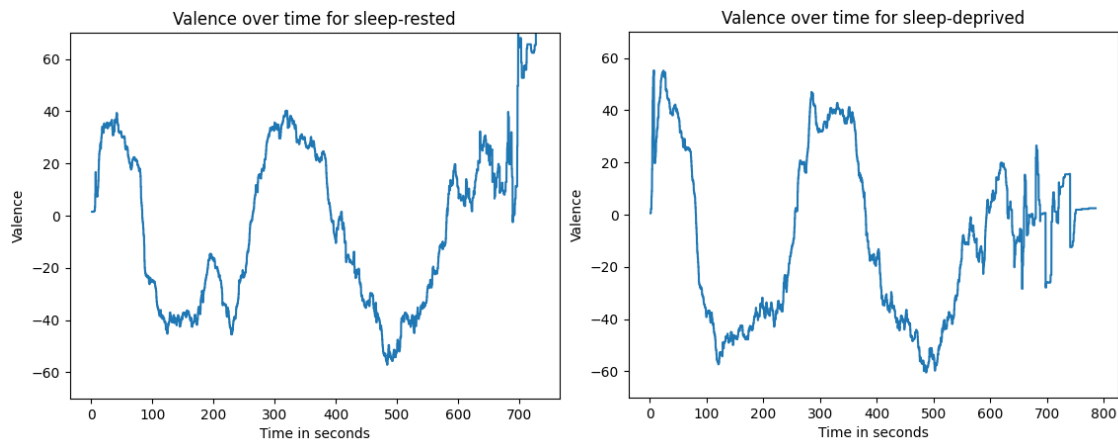


Figure 7.5: Illustration of mean subjective valence over time of the sleep-rested group (left) vs. sleep-deprived group (right) for the duration of the Underwood project experience sampled at 1Hz.

Results of the mixed linear regression model can be found in **Table 7.3**. The final formula of the model was:

$$v = g + P1 + P2 + P3 + P4 + P1 * g + P2 * g + P3 * g + P4 * g + (1|ID) + (1|r)$$

with v as the outcome representing subjective valence, g representing the group condition (sleep rest vs. sleep deprivation), $P1 - P4$ representing the four included principal components, $(1|ID)$ representing the random intercept of the participant ID and therefore addressing within differences between individuals, and $(1|r)$ representing the random slope per participant for each individual room. The model used 487 observations and 54 groups (one group per participant). The model was constructed using Restricted Maximum Likelihood (REML) estimation method. Significant main effects were found for group ($z = 2.57$; $p = .01$) and for PC2 ($z = -5.42$; $p < .001$), which represents the original predictors right-hand rotation distance and head rotation distance. A higher amount of rotating movement was therefore indicative of a valence decrease. Significant interaction terms were found too for the interaction of condition and PC1 ($z = 4.86$; $p < .001$); as well as the interaction of condition and PC2 ($z = 7.33$; $p < .001$). PC1 represents positional data, most notably right-hand position, left-hand position, head position, and player position (see **Fig 7.3** for more details). For the sleep-deprived group, this positional data positively predicted

valence, while for the sleep-rested group, it negatively predicted valence. An illustration of the observed significant interaction effects can be viewed in **Fig 7.6**).

Predictor	β	SE	z	p
Group	9.39	3.66	2.57	.01
PC1	-0.01	0.03	-0.15	.88
PC2	-0.38	0.07	-5.42	< .01
PC3	4.43	4.76	0.93	.35
PC4	-5.53	5.79	-0.95	.34
PC1 : Group	0.05	0.01	4.86	< .01
PC2 : Group	0.31	0.04	7.33	< .01
PC3 : Group	2.21	3.15	0.70	.48
PC4 : Group	4.07	3.62	1.13	.26

Table 7.3: Results of the mixed linear regression model.

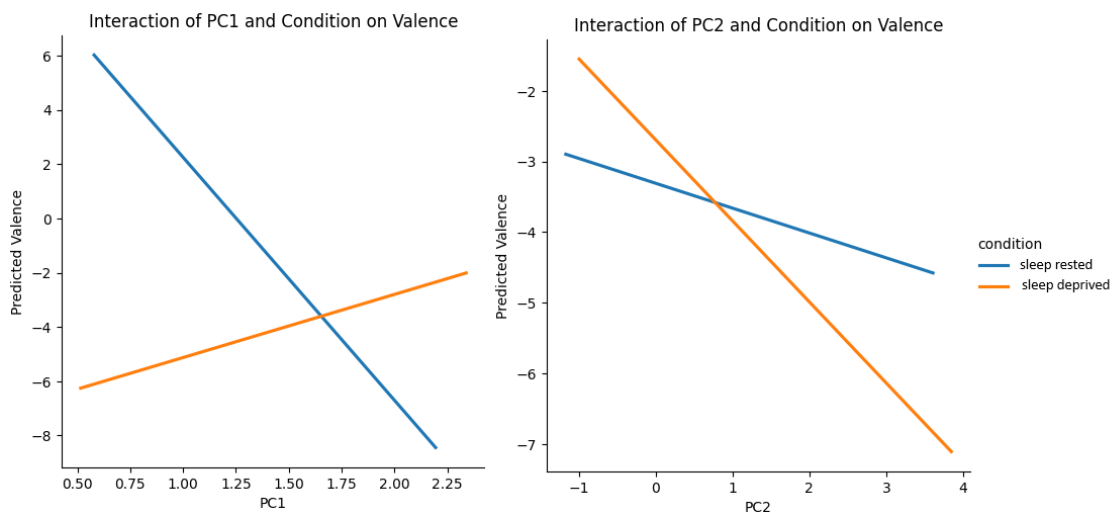


Figure 7.6: Illustration of interaction effects between condition and PC1 as well as condition and PC2 on predicted valence. The valence prediction made by the components changes significantly depending on the group condition (sleep rested vs. sleep deprived).

7.4 Discussion

This study tested whether or not specific contextual player characteristics (in this case sleep deprivation) could impact prediction models of emotional valence using a variety of predictors including physiological (EDA, HR) and position data during a fear-inducing virtual reality experience. Participants were randomly allocated into a sleep-deprivation and sleep-rest group and played through a custom-made VR horror experience. During

the experience, physiological and position data were measured and afterwards, participants rated their subjective valence during a screening of their playthrough. Using a mixed linear regression model, predictors and their interaction with the group condition were tested in their ability to predict subjective valence. Both the group condition and one principal component representing positions significantly predicted valence in this model. Two interactions between position data and group condition also showed significant effects, indicating that the prediction of affective journeys within game experiences is at least partly dependent on the amount of sleep players had the night before.

Building upon the findings of Sullivan et al. [343], sleep deprivation accompanies varying patterns of emotion expression and affective perception. Sleep deprivation has been shown to impair subjective recovery of anxiety and elevated physiological responses to ambiguous threats [346]. It can also contribute to the amplification of the anticipation of affective reactions [341] and lead to more negative classifications of neutral stimuli [359]. These findings all show that the lack of sleep as a condition of humans has the capability to alter emotion processing, potentially because of cognitive and neurophysiological patterns, such as difficulty in shifting attention [359] or in regulating parasympathetic activity in the brain [341]. In a wider context, these findings contribute to our knowledge relating sleep deprivation and anxiety, showing that anxiety expression - specifically through position - is influenced by whether or not participants have had a full night's sleep. Together with the findings by Sullivan [346], these results could provide guidance when it comes to identifying anxiety-related patterns and informing on potential prevention methods related to sleep.

Quality of sleep is doubtless an important factor when it comes to mental processes [341]. Given that it can be seen as a contextual variable that may or may not be expressed differently between individuals and for the same individual at different time points, this finding also has implications for our understanding of affective measures. If the positional patterns of a person relate differently to their anxiety level depending on their sleep quality, we cannot assume that position, in particular, can be used to measure emotions universally. This holds true for position measures but might also extend to other forms of emotional data, including biofeedback [32]. Our model did not show significant interactions between biofeedback components and the group condition in predicting emotional valence, meaning that we did not find evidence of changing the predictive power of physiological data depending on sleep deprivation. In fact, our model showed no significant effect for valence prediction for any physiological component, showing that HR and EDA cannot be assumed

to universally and sufficiently predict subjective affect in a virtual environment. While the position data contributed to valence prediction, the interaction with the group condition shows that this prediction cannot be universally assumed, since it depends on circumstantial characteristics of the participants (i.e. their sleep rest). For sleep-deprived participants, PC1 had a positive association with valence, while for sleep-rested participants the association was negative. This could indicate that affective expression in movement differs depending on the amount of sleep [346]. The interaction found with PC2 and condition also indicates that this could be the case, although here the association was not reversed, but merely steeper for sleep-deprived participants. Taken together, these results show that we cannot rely on "objective" or "universal" measures to model affective journeys. Emotion formation and expression is a complex multi-componential process where influences of context and individual differences need to be taken into consideration [2], [28].

From a practical standpoint, this has implications for all affective experiences, including video games. Creating an emotion-adaptive game is difficult [1], especially if the statistical relationship between affective measures and subjective emotion is unclear and dependent on external factors. Viewing potential measures with the lens of the Emotion Design framework (see **Chapter 4**) can be helpful in understanding how different measures relate to actual expressed affect. The concept of emotional input as described in **Chapter 4.2.3** proposed that player data can be used to infer emotional states if they are validated for a given player cohort, context, and game. According to this process, affective models can be created, but need to be closely related to the aims and the audience of a given game. My findings suggest that it is not enough to just measure heart rate or skin conductance and assume anxiety or hedonic levels based on this data. Multiple factors need to be taken into account when measuring and interpreting player data and this data needs to be validated for a given project to provide the most benefits. Predicting player anxiety in a virtual world is possible, but the process needs to adhere to methodological standards that are robust and take complex psychological interactions into account. In summary, these findings indicate that emotion prediction in a game context is not a straightforward process reinforcing the idea that no universal and objective emotion measure exists [32]. Even circumstantial player characteristics can systematically influence the statistical relationship between affective instruments and measured emotion. The concept of emotional input can help explain these effects and address potential solutions in real affective game implementations.

7.5 Conclusion

Just like the concept of emotional output, the concept of emotional input provides a way to conceptualize the difficult process of emotion measurement. We cannot just measure heart rate in a horror VR world and relate it to anxiety levels as the statistical relationship between measures and experience differs based on the context players are in. The concept of emotional input as well as the concept of emotional output can be seen as a useful tool to research and understand wide-ranging implications of affective human-computer interactions. Based on the presented findings here, the Emotion Design framework addresses an important gap in the field that has been observed in Chapter 3.3: Given the lack of universal mappings between games and player emotions or player emotions and measurement instruments, a strong methodological process for evaluating affective games is necessary. The work presented here aims to demonstrate that it is possible to study and develop affective games, but only if the process is embedded in a methodology that conceptualizes player emotions in all their complexity.

Chapter 8

The Future of Affective Systems

"It takes more than one human brain to create a human mind."

— Lisa Feldman Barret, *How emotions are made: The secret life of the brain* [38]

Up to this point, this thesis revolved around explaining the emotional player-game loop and bringing together concepts from psychology, design, and affective computing to create a robust framework to develop and evaluate affective games. This chapter will look at some potential future iterations of such systems. With progressing technologies, new applications are now possible that could combine the necessary functionalities of an affective game system as defined by Hudlicke [142] and Yannakakis [47]. Large language models (LLMs) specifically are theorized to have the potential to measure and adapt to emotional information communicated semantically. They have already demonstrated a potential to simulate a range of cognitive abilities [360] and even imputing a range of mental and affective states to others [361], without the need to fully understand the precise nature of the underlying cognitive processes. Because LLMs are trained on large text bodies that hold representations of human mental abilities, they have been observed to exhibit human-like performance on a variety of tasks [360], [362]. Since emotions are an important part of how humans perceive reality and therefore construct language [38], and are further heavily influenced by cognitive processes [36] including linguistic labelling [363], [364], language-based emotion representations might too enable deep learning models to understand and even simulate affective responses. They therefore could be used within affective systems to function as a complete semantic emotion measurement system, adaptation system, and elicitation system. As of yet, the potential of LLMs to fulfil these roles is however not well

understood.

This chapter aims towards clearing some open questions regarding the utility of LLMs working as an affective game system. Using contemporary findings of emotion research, a cognitive appraisal-based approach for language model affect generation is proposed and tested against other strategies to assess the ability to generate appropriate situational emotions. The results are used to implement affective agents within a newly developed conversational video game. The aims of this chapter are two-fold: (1) test if LLMs could be a useful tool to understand and simulate emotions as an affective game system; (2) build and evaluate an architecture that facilitates the effectiveness of emotion simulation as part of such systems. Different to traditional affect simulation agents, LLMs could make use of implicit information of large text bodies, even without a computational representation of complex mental abilities that are not fully understood. They could therefore improve on prior systems and lead to more immersive and more engaging virtual agents [39], [171].

8.1 Related Work

8.1.1 Affective Agents

In affective computing, researchers and developers are interested in creating affective systems that intelligently respond to changes in users' emotions [42]. Some of the benefits associated with affective computing techniques applied to video games include a more consistent, accessible player experience for a range of different players [46], personalized health and training applications [151], as well as new and purposefully designed gameplay mechanisms aimed at reinforcing target experiences [41], [47]. The use of affective agents in video games has been researched with special regard to this last aim. In 2011, Hudlicka discussed potential system design elements for affective (or more precisely, emotional) game agents [39]. According to the author, affective agents can be seen as computational representations of operationalizations made from emotion-theoretical models with appraisal functionality for emotion generation. For example, artificial agents could implement computational calculations of certain events to assess the relevance to the agent and consequently probable emotional reactions [39]. "Computation-friendly" appraisal implementations have often been built on models such as the OCC model [365] (see for example GAMYGDALA [173]). Taking specific fixed aspects (such as expectations of the agent [366]) into account, such models have been used to simulate appraisal based on decision trees.

The principal aims in regard to such artificial agents are considered natural, human-like behaviour and believability as part of a more fleshed-out and engaging game world [39], [41]. The tasks of agents therefore differ from other affective game mechanisms that mostly try to adapt the game world to player affect [158]. Procedural content generation (PCG) based on affective information has been shown to successfully increase enjoyment and offer personalized, immersive experiences in video games (see for example the work of Shaker et al. [231], [265]). This is often done by fine-tuning certain mechanics shown to be associated with a target player's emotion to increase the probability of that emotion [158]. Affective agents however do not need to adapt behaviours to player emotions but rather need their own emotion representations (or other natural representations of emotion components, such as simulated feeling or simulated physiology [39]) that could then lead to believable behaviours.

The central issue of designing and developing affective game agents lies therefore in creating good computational simulations of emotional states. Human emotions are complex psycho-physiological states that are expressed within behavioural, physiological, cognitive, and feeling components [28]. Moreover, while much work has been done to empirically investigate emotions, many core theoretical disagreements remain [28], including debates between dimensional [61], discrete [79], constructivist [38], and cognitive [55] perspectives.

A fully developed affective agent would make it necessary to first solve all fundamental psychological gaps that have been present since the beginning of affective computing [42], and then integrate them into working, computational systems [39], [47]. This means that building a psychology-based, fully functional, and accurate emotion simulation for an artificial agent is currently not possible and would be in almost all game design cases impractical. However, we may still be able to build affective agents that possess key features of emotion elicitation in humans and, as a consequence, allow for relatively successful simulation of human emotions. One candidate feature is appraisal. Emotion elicitation is dependent on contextual and individual factors, processed through appraisal [36], [88], [367]. The notion of emotion appraisal is that emotions are caused by subjective evaluations of triggering events regarding their significance to one's personal life or interests [89]. Evidence suggests that appraisal holds a central role in emotion elicitation and as a consequence acts on all other emotion components [367].

Any given external (e.g., situations) or internal (e.g., thoughts) event may be appraised on multiple variables that contribute to emotion formation. Such variables might include

goal relevance, certainty, coping potential, or agency [36]. Appraisal therefore represents a flexible process that adapts to individual differences [368] and the current context [106]. Evidence also suggests that language can play a key role in emotional appraisal, both by providing key contextual information from which to construct the appraisal and by providing conceptual labels for the appraised states [364]. With all this in mind, language models might provide one means of simulating the appraisal process as they can generate high-level meaning-driven text outputs based on text training data that potentially holds implicit representations of human psychological processes [360].

8.1.2 Language Model Approach

In the last few years, Natural Language Processing (NLP) has been rapidly progressing to the point that single task-agnostic language models perform well in a range of tasks [369], including the simulation of human-like behaviour [362]. The basis for this is the large amount of training data representing a wide range of human behaviour through language [369], [370]. Concerning games, models such as OpenAI's Generative Pre-trained Transformer (GPT-2 and its successors) have shown early successes in the procedural generation of interactive stories [371], text-based adventure dialogue as well as action candidates [372], [373]. In a recent simulation study by Park et al. [374], language models were implemented in artificial agent architectures to populate a sandbox world reminiscent of *The Sims* [375]. The architecture includes storing and retrieving information from a memory system based on relevancy for the current situation and then uses the information to generate reflections (i.e., high-level interpretations of a situation), plans (i.e., potential future actions), and immediate actions. Multiple agents were simulated in a game-like world and the authors suggest that emerging interactions were natural and believable in terms of human-like simulation.

While work in this area is still in an early stage, the use of language models addresses some concerns with prior approaches. Most notably, instead of trying to build computational representations of human behaviour, the main task involves trying to retrieve believable human behaviour given a situation from a language model and implementing the results within a game agent. Depending on the game aim, this involves (1) translating the current situation with regards to the expected outcome into language; (2) generating content using a large language model; and (3) translating the output back in order to implement it in a game system. For example, Ciolino et al. [376] used a fine-tuned

GPT-2 model to generate Go moves by translating the current board state to text and the language output back to action suggestions. Such a process is naturally easier for purely text-based tasks, such as dialogue generation, where text is already the expected output and the expected output can comparatively easily be described in language [372], [374].

Still, even purely text-based generative tasks can pose some potential barriers for language models. The most obvious barrier comes from the underlying training data. No language model can represent human behaviour in its entirety but is limited to the training data and its biases [377] as well as model specifications [369]. Additionally, the performance of a language model is not only dependent on the output but also on the input [378]. For example, chain-of-thought prompting is a concept introduced by Wei et al. from the Google Research team [379] and relates to the integration of chain-of-thought in few-shot prompts to improve the reasoning capabilities of language models. Similarly, as Park et al. [374] describe, simulating believable behaviour in a digital world includes various important steps (including storing and retrieving memory, reflecting in addition to observing, etc.) that ultimately work together to improve the probability of generating expected and natural behaviour.

When it comes to designing affective agents (i.e., agents that simulate emotions), the first question that we have to ask is how well affect is represented in the training data and whether a language model is capable of retrieving it. In a recent paper discussing the performance of different GPT iterations on theory-of-mind tasks, Kosinski [361] found that new language models perform very well when it comes to imputing unobservable mental and affective states to others. Such findings (especially combined with findings indicating good performance on cognitive tasks [360], [380]) suggest that high-level psychological mechanisms are represented in language alone and could therefore be simulated with a well-constructed language model. Along these lines, can we effectively and efficiently achieve accurate and natural affect-simulation using language models? If we can assume that emotions are represented in language models, mechanisms for emotion elicitation (such as appraisal) might also be represented. And given that language models can be improved through in-context learning [369], [378], for example by chain-of-thought prompting [379], affect generation might be facilitated by architectures that allow for affective in-context learning. This study therefore discusses the potential of language models to simulate affective game agents by testing affect generation capabilities of different implementation architectures, including a newly developed appraisal-based architecture to

facilitate a natural chain of emotion.

8.1.3 Appraisal-Prompting Strategy for Emotion Simulation

The overall aim of this study is to create an effective affective agent architecture for a conversational game (i.e., a game with language-based user input and language-based agent response). However, since language models have also been successfully used to generate agent action spaces [372], [374], this process could also apply to other simulations of human-like affect in non-playable video game characters.

The basis of the approach is rooted in traditional PCG research, particularly in studies integrating affect-adaptation (see for example [158], [231]). User input (which in this case is text input) typically gets parsed into the game logic to adapt the content in a meaningful way, for example, to better elicit a target experience [47]. This could mean that game agents react to certain player inputs or even their own interactions in the game world [374]. To make use of language model functionality, interactions need to be translated into language. Depending on the specific language model in use, the language should follow specific patterns to yield the best result, which is generally known as prompt engineering [381]. One pattern that can be considered inherently relevant to simulating game agents is persona patterns, which instruct an LLM to assume a certain characterized role. Combined with aspects of game-play patterns that provide the abstract game context for the persona tasks [381], the most basic form of an interaction synthesized with such patterns solely includes player-provided text that is used to generate responses. Because emotions are represented in language models [361], this very basic step alone could make a rudimentary affective agent.

However, static prompt patterns have limitations for creating believable game agents. Most notably, they do not incorporate memory as a basic foundation of human interaction. Applications that integrate language models, such as ChatGPT, partially address this by logging an input-response history that influences progressive content generations, which can create more natural conversation flows and improve the performance of future generations [369], [379]. In its most basic form, memory could integrate the preceding in-game observations (such as the course of a player-agent dialogue) into the following prompts. In other words, it expands the prompt pattern to include memorized in-game observations to facilitate in-context learning [378]. This has however two major constraints: First, prompts are limited in terms of possible length and complexity, meaning that a full characterisation

of a game agent cannot be included in a prompt for every given generation. Park et al. [374] addressed this problem by designing a memory system that first retrieves relevant entries from a data base before constructing prompts for the language model. Another less resource-intensive solution for simpler simulations could be to only store new information if it is considered relevant for future generations. A second limitation is a lack of deeper understanding. Tracking only observations makes it hard for a language model to draw inferences or make generalized interpretations of a scenario [360]. This problem could be addressed by introducing other types of memories, such as reflections and plans [374]. For affective agents, the more appropriate information to track in addition to external observations would be their internal emotional progression - or chain of emotion.

To summarize, language-based affective game agents need some kind of memory system in place that stores observations and emotions. This memory system is the base for future prompts. For simple games, such as the short conversational game developed for this study, only relevant information is stored in memory, which replaces a retrieval system as game agents have a limited pool of expected actions that can be considered at the time of memory storage. More complex games that simulate agent behaviour should, however, consider a memory retrieval system instead [374].

To store emotions and therefore create a chain of emotions, the architecture needs a system that turns observations into emotional reactions. Because emotion elicitation is highly dependent on appraisal with consideration of the current context and individual differences [36], [367], this system could make use of appraisal prompting, i.e., the use of contextual information and characterizations for the agent to appraise a current situation to generate current emotions. As shown in Fig 8.1, initial context information and character information can be provided by the game designer and stored in the memory system of the affective agent. The appraisal system would then expand the stored memories to include current emotions for every observed behavior and therefore create a chain of emotions. This, in turn, could be used to generate the agent's behaviour (specifically in terms of a conversational game, the agent's dialogue).

One aspect to consider when developing affective agents is evaluation criteria. Different to most cognitive abilities [360], there are few standardized benchmarks for successful emotion simulation. In psychology, one indicator of the ability to appraise and express emotions is Emotional Intelligence (EI) [382]. EI is considered an ability, influenced but not dictated by cognition [383], and is therefore often used to assess the emotional capabilities

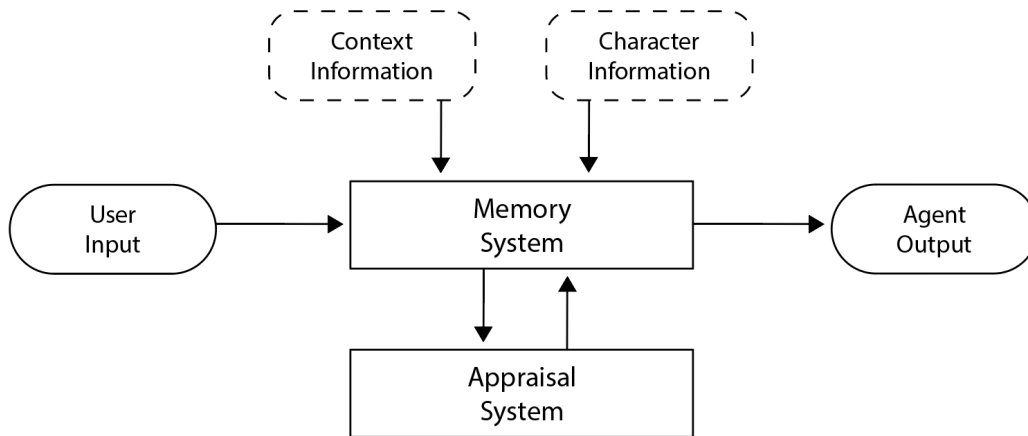


Figure 8.1: Schematic representation of the proposed architecture. Agents are set up by providing relevant context and role-playing character information already integrated into a memory system. In interactions, user input gets stored in the memory system and triggers appraisal (i.e., explicit emotion expression) that is also stored in the memory system. Based on the current state of the memory system, agent output is generated.

of children and adults in various settings [384]. However, the definition of EI as a construct is fuzzy and many instruments are criticized for measuring potentially different abilities all under the umbrella term EI [385]. As a consequence, MacCann and Roberts [386] developed new measures for more precisely defined dimensions, including the Situational Test of Emotional Understanding (STEU), which relates to the ability to appraise the appropriate emotion for a given situation.

While emotional understanding can be argued to be the central ability an affective agent must have, in the context of affective games, user experience becomes much more relevant. For example, one central aim of game agents is to display believable and human-like behaviour [366] or personality and social presence [387]. The ability to understand and create more accurate emotions is therefore only one aspect to consider when evaluating the success of affect simulation in video games and more user-centered methods need to be investigated as well. The proposed architecture will therefore be tested on multiple domains, including emotional understanding, agent believability, as well as user-perceived emotional intelligence, warmth, and competence. We present three distinct experiments to achieve this: The first experiment tests a common LLM in terms of emotional intelligence

with the validated STEU measure to assess how well situational emotional reactions can be simulated. The second experiment compares the proposed architecture to two control architectures in terms of content generation of an emotional situation that is qualitatively and quantitatively analysed. The final experiment consists of an implementation of the proposed architecture within a conversational video game to test against the same control architectures in terms of user experience in a randomized user study. Together these three experiments are used to shed light on the feasibility of using LLMs to simulate affective agents and evaluate emotional LLM architectures on multiple objective and subjective measures.

8.2 Experiment 1: Investigating Situational Emotional Understanding Using Appraisal-Prompting

8.2.1 Materials and Methods

To assess the capabilities of a language model in appraising emotions in various situations, this first experiment implements the language model GPT-3.5-turbo by OpenAI (accessed through the API) [388] to answer the 42 items of the STEU [386]. Each STEU item presents a situation (e.g., "Clara receives a gift.") and a question with five possible answers, one of which is correct (e.g., "Clara is most likely to feel? [A] Happy [B] Angry [C] Frightened [D] Bored [E] Hungry").

All items were answered three separate times, involving three prompting strategies: The first strategy represents the baseline capabilities of the model to appraise human emotions, as it only reflects the model's outputs when prompted with each STEU item separately presented together with the example item and example answer. The second strategy implements memory and therefore context-based learning, as all prior items and answers are included in subsequent prompts. The third strategy expands this process by changing the answer of the example item to a 2-step answer: First, the situation is appraised based on the contextual information to provide the most likely emotion and in a second step, the item is answered. This last strategy therefore tests if the implementation of appraisal in prompting yields better results for emotional appraisal. Figure 8.2 shows the input and output for the first STEU item, including the example item for No-Memory/Memory (as the input is the same for the first item for these two conditions) vs. Appraisal Prompts. Consecutive input consisted of the next STEU item and included either again the example

item (for the No-Memory condition) or all previously answered items and responses (for the Memory and Appraisal Prompting conditions).

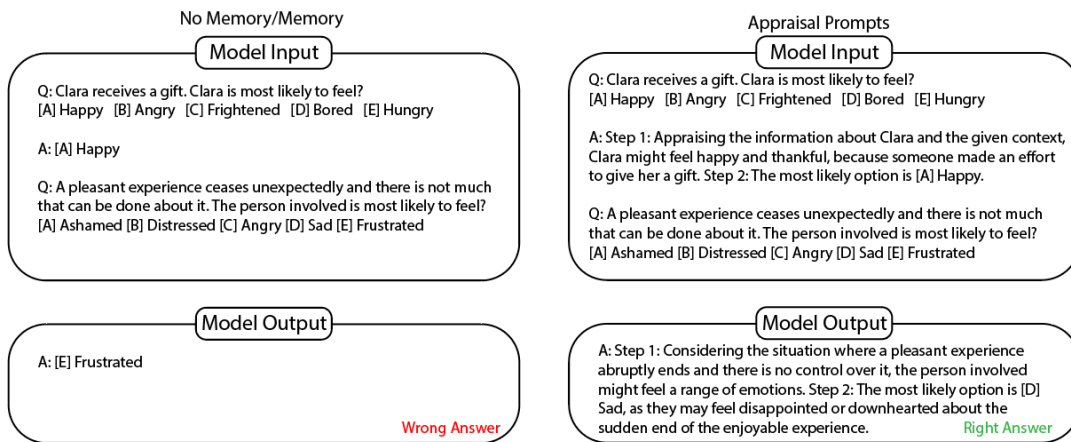


Figure 8.2: Example of model input and output for the three conditions. The input of the No-Memory and Memory condition is the same for the first item. For the No-Memory condition, all following items only include the example question and answer, as well as the next question in the scale. The Memory condition includes all prior questions and generated answers. The Appraisal-Prompts condition is the same as the Memory condition, but the example answer is changed to include two steps: First, appraising the situation to generate emotions of the involved person and second, providing the answer.

Similar to the process shown by Binz et al. [360], default values were kept for all parameters, except temperature, which was set to 0 to ensure deterministic responses.

8.2.2 Results

The language model was able to solve the tasks presented within the STEU in each condition above the chance level. In the No-Memory condition, the language model was able to successfully solve 24 out of 42 items, which represents a mean score of 0.57 that was noticeably higher than chance level (0.20). In the Memory condition, the language model solved 31 out of 42 items, which represents a mean score of 0.74. In the Appraisal-Prompts condition, the model was able to solve 35 out of 42 items, which is a score of 0.83 and therefore represented the best performance of all conditions. Table 8.1 shows a summary of the descriptive statistics for all three conditions. Figure 8.3 displays the results of each condition for each STEU item. As the figure shows, the Appraisal Prompts perform ei-

ther as well as the other conditions or outperform them on all STEU items. One notable exception was item 30 ("An upcoming event might have bad consequences. Nothing much can be done to alter this. The person involved would be most likely to feel? [A] Sad [B] Irritated [C] Distressed [D] Scared [E] Hopeful"), which was only correctly solved by the Memory condition (with D as the correct answer compared to C, which was chosen by the other conditions).

Condition	Sum	M	SD
No Memory	24	0.57	0.50
Memory	31	0.74	0.45
Appraisal Prompts	35	0.83	0.38

Table 8.1: STEU scores (out of 42) by condition. Each STEU item can either be right (1) or wrong (0)

8.3 Experiment 2: Content of an Appraisal-Based Chain-of-Emotion Architecture

Given the potential found in the previous experiment, the next logical step is to implement the appraisal-based strategies into a role-playing agent architecture and compare the results with control architectures on various outcomes. The following section describes a mixed-methods approach to evaluate the success of each implemented architecture within a conversational role-playing scenario. This study uses fixed prompts and compares outputs of three different conditions in terms of their emotional content.

8.3.1 Materials and Methods

8.3.1.1 Scenario

To test the different strategies within real game architectures, a role-playing scenario was introduced. The setting for this scenario was a cafe called "Wunderbar", where the language model was tasked to play a role-playing character (called "Chibitea") meeting their long-term romantic partner who requested the meeting to ultimately break up. This scenario was chosen because of the depths of possible emotional responses from the agent and created through simple conversational exchanges. The instruction prompts and the fixed inputs used for all conditions can be found in the Appendix A.3. The agent's responses were again generated using OpenAI's GPT-3.5-turbo model (accessed through the API)

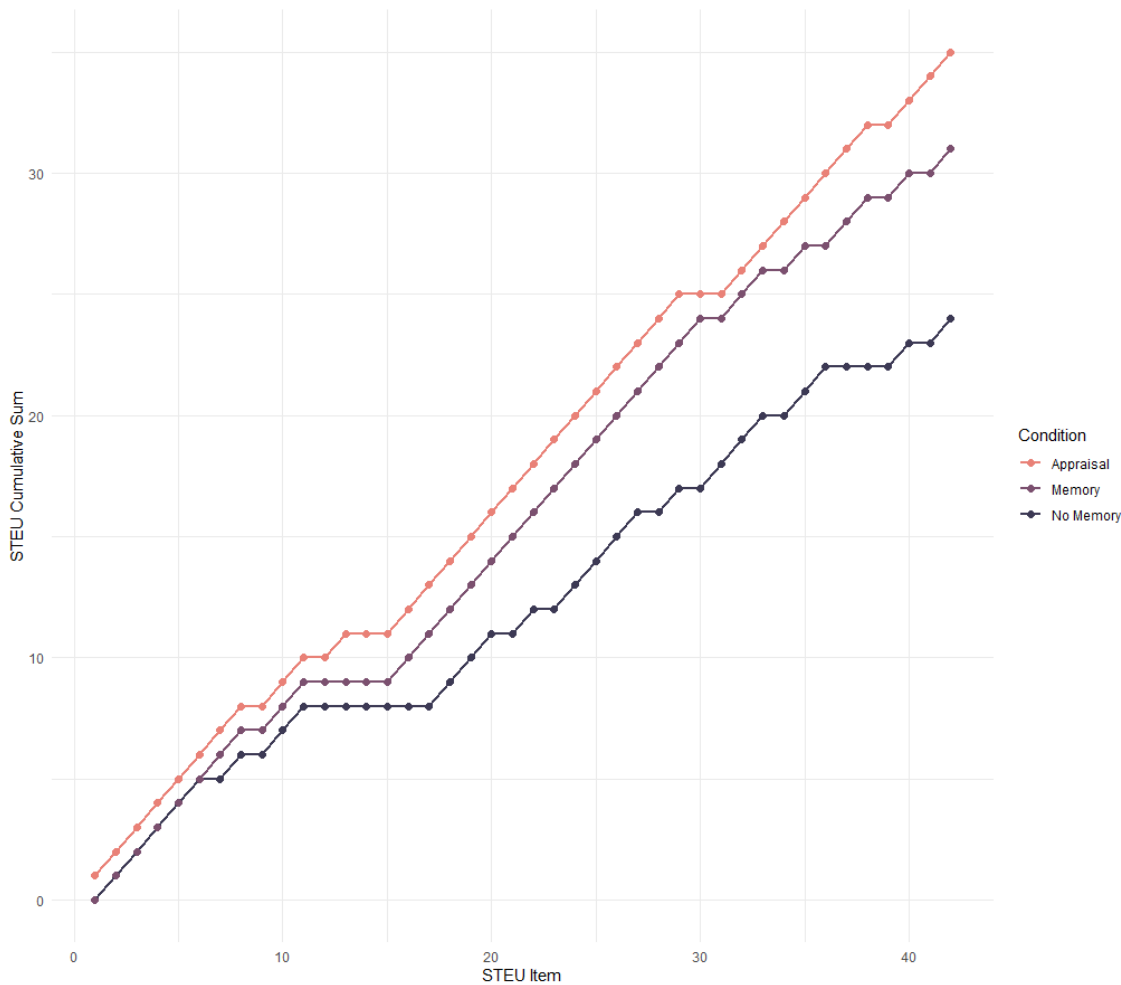


Figure 8.3: Results of the comparison between conditions. The Y-axis represents cumulative STEU score by item and the X-axis represents individual STEU items.

[388]. All LLM parameters were kept to their default values, except temperature, which was set to 0 to ensure deterministic responses.

8.3.1.2 Conditions

Again, three strategies of emotion generation were compared. The No-Memory condition can again be seen as a baseline/control condition for the system’s ability to simulate appropriate emotional responses from the fixed inputs and task instruction. All details about the agent’s character’s personality and important context that would facilitate appraisal were given in the task description within the No-Memory condition.

The Memory condition stores each user input and generated response as a memory data structure that keeps the entire conversation log in memory. Because the tested scenario was

rather short, it was possible to keep the memory fully in the context window to create new generations, making retrieval functionalities unnecessary. This system therefore represents a prompt construction system involving the task instruction and prior conversation log.

The Chain-of-Emotion condition implemented the appraisal model shown in Figure 1 and involved two steps: First, appraisal prompting (see Experiment 1) is used to generate the current emotion of the agent, which is then in the second step implemented into the prompt for response generation. For the first step, appraisal prompting was achieved with the following prompt: "Briefly describe how Chibitea feels right now given the situation and their personality. Describe why they feel a certain way. Chibitea feels:". The generated text-based emotion descriptions were stored in the memory system and represented a chain of emotions of the agent for the duration of the game. For the second step, again the entire chat history was included in the prompt, but this time included the generated emotions from the first step. This condition therefore represents a 2-step process of first generating a fitting emotion of the agent using appraisal prompting, and then generating a text response similar to the Memory condition, but with the addition of the stored chain of emotion. Figure 8.4 illustrates the prompting strategies for each condition. The complete prompts for this experiment can be found in Appendix A.3.

8.3.1.3 Measure

Fixed inputs were used to create responses from each implemented agent architecture, which were analyzed in terms of their emotional content, using the Linguistic Inquiry and Word Count tool (LIWC; [389]), a content analysis tool based on word occurrences often used in affective computing [42] and psychology studies [390] to analyze emotion expression. The tool provides a word count for each text segment (e.g., per sentence), a proportion of affective words (% of affective words per sentence), as well as on a more detailed level a proportion of positive emotion words (% of positive affective words per sentence) and negative emotion words (% of negative affective words per sentence). Finally, the LIWC also calculates scores for authenticity (see [391] for details) and emotional tone, which signalizes the proportion of positive words compared to negative and neutral words (see [392] for details).

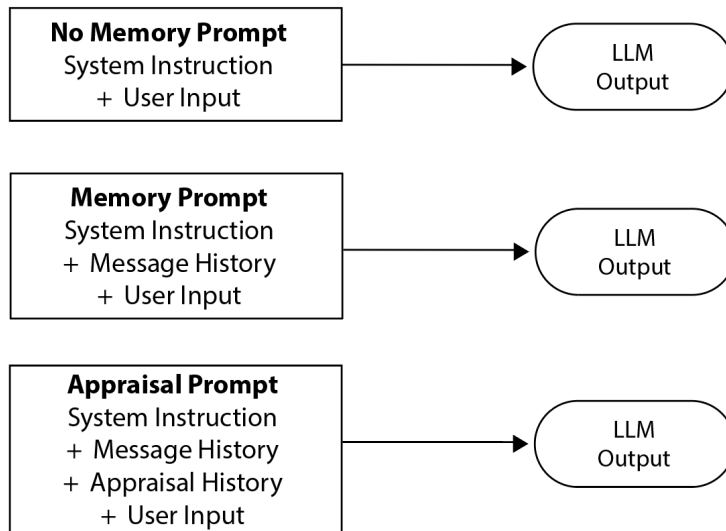


Figure 8.4: Illustration of prompting strategies for all three conditions. The No-Memory condition constructs a prompt out of the system instruction and the user input. The Memory condition constructs a prompt out of the system instruction, message history, and user input. The Chain-of-Emotion condition uses a separate LLM call to appraise the agent’s emotion for each message before creating a user response. It therefore constructs a prompt out of the system instruction, the message history, the history of the emotions generated through the appraisal step, and the user input.

8.3.1.4 Procedure

Pre-written prompts were used that stayed constant between all conditions in order to gauge the qualitative characteristics of each condition response. A list of the resulting conversations within all three architectures can be found in the Supporting Information A.3. The generated content was qualitatively described and the LIWC was used to analyze the content quantitatively. To achieve this, the generated output was separated into individual sentences and mean scores were calculated for each measure of interest (see Table 8.2).

8.3.2 Results

When analyzing the descriptive attributes of each text (as a common content analysis approach [393]), we can observe that the Chain-of-Emotion condition initially generated more specific memories for the time with the player (“Remember that time we got lost in the enchanted forest and ended up finding that hidden waterfall?” as opposed to “Re-

member all the adventures we've had together?"). For the duration of the conversation, the emotional journeys of all three conditions began to diverge. For example, because the No-Memory system had no recollection of previous exchanges, the overall emotional expressions remained in a state of anticipation ("I feel a mix of excitement and anticipation"). The Memory system showed a progression, starting from expressions of love and happiness to shock, confusion, sadness, and fear to finally expressions of hope ("I hope we can find happiness, whether it's together or apart."). The Chain-of-Emotion system showed indications of complex mixed emotions even very early in the conversation ("What matters most to me is your happiness, even if it means letting go") as opposed to the pure expressions of pain and sadness in the other conditions. This continued when the system was prompted about past memories ("I feel a mix of nostalgia and gratitude" as opposed to "I feel an overwhelming sense of love, joy, and gratitude" in the Memory condition). The Chain-of-Emotion condition also used more implicit affective expressions ("I... I never expected this" as opposed to "I'm shocked and hurt" in the Memory condition; "There is so much I want to say but words fail me in this moment" as opposed to "I want you to know that I love you deeply" in the Memory condition).

Using LIWC to make the text contents quantifiable, we observed significant differences in mean Authenticity score per sentence by condition ($F[1,71] = 5.10; p = 0.03$). Follow-up t-tests revealed significant differences between the Chain-of-Emotion condition and both the Memory condition ($t[34.3] = -2.29; p = .03$) and the No-Memory condition ($t[31.1] = -2.30; p = .03$). Descriptive statistics of all tested LIWC variables can be found in Table 2.

8.4 Experiment 3: User Evaluation of Game Implementations

In this study, users were asked to play through an interactive game version of the scenario introduced in Study 2 to evaluate each agent architecture for multiple outcomes (specifically agent believability, observed emotional intelligence, warmth, and competence). This study therefore expands on the findings of Experiment 2 by implementing the architectures and the scenario within a video game and evaluating all three conditions in terms of user experience measures.

	No Memory ($N = 22$)	Memory ($N = 24$)	Chain of Emotion ($N = 27$)	F (p)
	M (SD)	M (SD)	M (SD)	
Word Count	18.00 (6.92)	15.20 (4.73)	17.00 (7.59)	0.20 (.65)
Authentic Score	61.50 (38.60)	61.9 (39.50)	82.60 (21.20)	5.10 (.03)
Tone Score	74.20 (38.60)	62.00 (44.80)	53.90 (44.00)	2.76 (.10)
% Affective Words	11.40 (8.19)	13.50 (11.00)	10.65 (7.82)	0.08 (.78)
% Positive Emo- tion Words	5.28 (6.29)	4.13 (4.82)	3.32 (4.35)	1.76 (.19)
% Negative Emo- tion Words	0.59 (1.97)	3.10 (7.56)	1.57 (2.89)	0.37 (.55)

Table 8.2: Descriptive overview of LIWC variables per output sentence by condition for the fixed prompt responses with F and p values of the significance test.

8.4.1 Materials and Methods

8.4.1.1 Conversational Game

A conversational role-playing game was developed based on the scenario tested in Experiment 2. The setting of the game was again a cafe called "Wunderbar", where this time the role-playing character of the player (called "Player") requested to meet their long-term romantic partner (called "Chibitea"). The aim of the players was to play out a breakup scenario with the game agent within six interactions. The players' characters had the specific in-game aim of breaking up, while the agent's character had the aim of staying together.

Players were instructed to not worry about creativity but rather to stay in character for the interactions and be observant of the AI agent's emotional reactions. Players were also instructed to make up reasons for the breakup. In-game screenshots can be viewed in Figure 8.5. The agent's character is procedurally generated from different body parts and colour palettes, providing visual variation each time the game is played. To ensure that these generations had no systematic influence on player responses, the possibility space was made very large (5,184 different possible character designs).

The user interface was deliberately kept simple. For each playthrough, the agent would greet the player through a text bubble. Then the player was prompted to answer via a simple text input field. Player answers were submitted through a button next to the field. The input then disappeared until the agent's answer was rendered. The game ended

after 8 dialogue exchanges (i.e., 8 player messages and 8 agent responses). The game was developed using the Unity game engine with C# as a scripting language.

As with Experiment 2, the agent’s responses were generated using OpenAI’s GPT-3.5-turbo model (accessed through the API) [388]. The game also made use of the moderation API to test each generated response for harmful or inappropriate messages [394] that would end the game on detection of such messages. As with the previous studies, all LLM parameters were kept to their default values, except temperature, which was set to 0 to ensure deterministic responses.

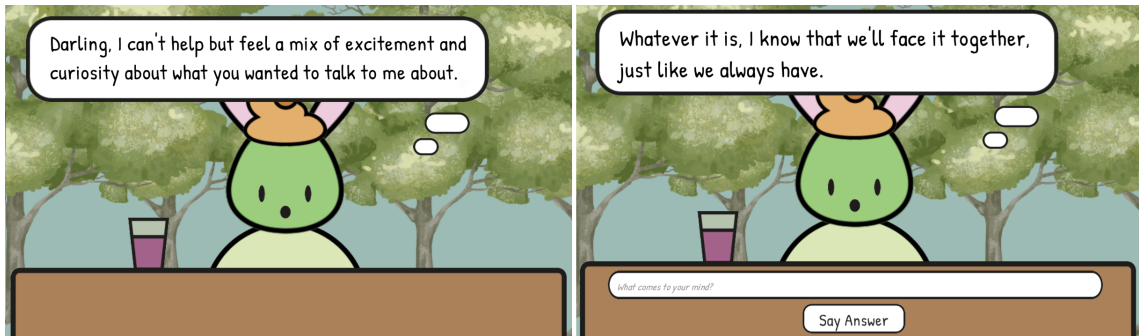


Figure 8.5: Screenshots of the conversational game “Wunderbar”. The left screenshot shows dialogue provided by the model. The user can click to continue each dialogue line until the input field for a response appears (right screenshot).

8.4.1.2 Conditions

The game implemented the same architectures used in Experiment 2. The No-Memory condition therefore represented generated LLM responses solely based on user input and instruction prompts. The Memory condition incorporated the conversation log into a memory system that was again short enough to be completely included in the prompts for the language model without a retrieval system. The Chain-of-Emotion system was also constructed exactly as in Experiment 2 and included the same instruction prompts, therefore again involving an initial appraisal step before responses for the agent were generated.

8.4.1.3 Measures

Players were asked to fill out three short questionnaires for each tested architecture. The first questionnaire was an adaptation of the four agent believability questions used by Bosse

and Zwanenburg [366]. The four items were “The behaviour of the agent was human-like”, “The agent’s reactions were natural”, “The agent reacted to my input”, and “The agent did not care about the scenario”. The second questionnaire comprised four items measuring observed ratings of emotion intelligence, adapted from Elfenbein et al. [395] who originally adapted these four items from the Wong and Law Emotional Intelligence Scale (WLEIS; [396]) by replacing the word “I” with “he/him” or “she/her”. For our study, we changed these words to ask about “the agent”: “The agent always knows their friends’ emotions from their behaviour”, “The agent is a good observer of others’ emotions”, “The agent is sensitive to the feelings and emotions of others”, “The agent has a good understanding of the emotions of people around them”. Finally, the third questionnaire measured players’ assessment of the agent’s personality along the two classic stereotype dimensions warmth and competence with two items each (warm and friendly; competent and capable) as described by Cuddy et al. [397]. In combination, these 12 items assessed players’ perception of the agent’s believability as a human-like character, the agent’s emotional intelligence, and the agent’s personality on the classic dimensions of warmth and competence.

8.4.1.4 Procedure

A pilot study was conducted before recruitment began. 5 participants (1 female) with a mean age of 27 played through the game once for each of the three conditions. After each version, they answered the 12 questions from the three included questionnaires. Following this, participants were asked for demographic data (age and gender) and the experiment ended. Feedback from all pilot participants was gathered and used to improve the consistency of the game and data logging implementation. The final study was then created as a WebGL build and made available online via the free video game hosting platform itch.io.

During the main experiment, participants were asked to carefully read the study information sheet and agree to participate voluntarily via the consent form. They were informed that participation was subject to OpenAI’s usage terms for prompt writing, while the GPT output was controlled through an implementation of OpenAI’s moderation API. Participants then progressed through the three game scenarios similarly to the pilot testers in a within-subject design. The presentation order of the conditions was counter-balanced between participants to ensure that no systematic order effects could influence results.

8.4.1.5 Participants and Statistical Analysis

A total of 30 participants (10 female) were recruited through the institutional subject pool of the authors. The recruitment period started on July 25th, 2023 and ended on August 14th, 2023. Participation was compensated with University credits if applicable. The sample size was considered appropriate based on a statistical power analysis, yielding a power of 0.95 for medium-sized effects (0.5 SD) in repeated measures ANOVAs. The age of participants ranged from 19 to 47 years ($M=26.41$; $SD=7.27$).

Within-subject ANOVAs were conducted for each measure (agent believability, observed EI, warmth, and competence). Follow-up t-tests were used to identify specific differences between conditions for each measure. All analyses were conducted in R. The underlying data was made available via the Open Science Framework [398]

8.4.1.6 Ethics Statement

Written consent was granted after reviewing the methods of our study by the Ethics Committee of the Psychology Department of the authors' institution. The experiment was conducted in accordance with the recommendations of this committee.

8.4.2 Results

Multiple significant effects between the three conditions were observed in the user study. An overview of descriptive statistics can be found in Table 8.3. First, an effect was found for the believability item "The agent's reactions were natural." ($F[2,84] = 3.65$; $p = 0.03$). Follow-up t-tests revealed differences between the No-Memory and Chain-of-Emotion condition ($t[41.46] = -2.79$; $p = .008$), as well as the Memory and Chain-of-Emotion condition ($t[52.47] = -2.00$; $p = .05$). There was also an effect for the believability item "The agent reacted to my input." ($F[2,84] = 3.62$; $p = 0.04$). T-test revealed that this effect was based on a difference between the No-Memory and the Chain-of-Emotion condition ($t[40.41] = -2.41$; $p = .02$).

Regarding the EI questions, an effect was found for the item "The agent is sensitive to the feelings and emotions of others." ($F[2,84] = 3.31$; $p = 0.04$). Follow-up t-tests revealed differences between the No-Memory and Chain-of-Emotion condition ($t[43.84] = -2.70$; $p = .01$), as well as the Memory and Chain-of-Emotion condition ($t[40.52] = -2.07$; $p = .04$). There was no statistically significant difference between the conditions when it comes to observed personality aspects.

	No Memory ($N = 30$) M (SD)	Memory ($N = 30$) M (SD)	Chain of Emotion ($N = 30$) M (SD)	F (p)
"The agent's behaviour was human-like."	4.82 (2.36)	5.36 (1.99)	5.75 (1.71)	1.95 (.15)
"The agent's reactions were natural."	4.43 (2.22)	4.89 (1.95)	5.71 (1.01)	3.65 (.03)
"The agent reacted to my input."	5.43 (2.20)	6.29 (1.38)	6.54 (1.04)	3.62 (.04)
"The agent did not care about the scenario."	3.32 (2.28)	3.14 (2.29)	3.32 (2.36)	0.26 (.78)
"The agent always knows their friends' emotions from their behaviour."	4.79 (2.10)	5.04 (2.03)	5.71 (1.46)	1.32 (.27)
"The agent is a good observer of others' emotions."	4.89 (2.20)	5.29 (2.21)	5.93 (0.90)	2.25 (.11)
"The agent is sensitive to the feelings and emotions of others."	5.14 (1.94)	5.39 (1.97)	6.25 (0.97)	3.31 (.04)
"The agent has a good understanding of the emotions of people around them."	4.86 (2.07)	5.11 (2.20)	5.61 (1.40)	0.86 (.43)
"How capable was the agent?"	4.86 (2.01)	5.57 (1.45)	5.14 (2.35)	1.68 (.19)
"How competent was the agent?"	5.39 (1.55)	5.39 (1.85)	5.11 (2.06)	1.68 (.19)
"How friendly was the agent?"	4.86 (2.05)	5.18 (1.89)	5.00 (2.13)	0.41 (.66)
"How warm was the agent?"	6.07 (1.07)	6.38 (0.73)	5.55 (2.11)	2.48 (.09)

Table 8.3: Descriptive overview of user research variables per condition with F and p values of the significance test. Each item has a minimum value of 0 and a maximum value of 6.

Again, the LIWC was used for content analysis of the generated texts. Significant differences in mean Tone Score by condition were observed ($F[1,574] = 12.28$; $p < 0.001$).

Follow-up t-tests revealed significant differences between the Chain-of-Emotion condition and both the Memory condition ($t[383.7] = 2.02$; $p = .03$) and the No-Memory condition ($t[374.94] = 3.53$; $p < .001$), as well as a difference between the Memory and No-Memory condition ($t[1367.6] = 4.09$; $p < .001$). A descriptive overview of all tested LIWC variables can be found in Table 8.4.

	No Memory ($N = 30$) M (SD)	Memory ($N = 30$) M (SD)	Chain of Emotion ($N = 30$) M (SD)	F (p)
Word Count	64.40 (26.60)	59.60 (21.20)	62.30 (34.10)	0.52 (.47)
Authentic Score	70.40 (29.30)	74.00 (27.60)	72.00 (24.80)	0.32 (.57)
Tone Score	84.40 (28.50)	80.50 (32.60)	73.00 (34.90)	12.28 (.00)
% Affective Words	9.76 (3.19)	10.10 (4.14)	9.61 (4.46)	0.15 (.70)
% Positive Emotion Words	3.84 (2.35)	3.78 (2.62)	3.64 (2.72)	0.60 (.44)
% Negative Emotion Words	0.68 (1.35)	0.69 (1.48)	0.93 (1.35)	2.76 (.10)

Table 8.4: Descriptive overview of LIWC variables per participant by condition for all outputs generated in the user study with F and p -value of the significance test.

8.5 Discussion

This study investigated the emotional intelligence capabilities of LLMs using different prompting strategies (No-Memory, Memory, Appraisal) and found better performance for appraisal-prompting strategies when it comes to successfully identifying adequate emotions in different theoretical situations (Experiment 1). These findings were then used to create a Chain-of-Emotion architecture for affective game agents that was tested in a role-playing scenario in terms of content output against traditional LLM architectures (Experiment 2). Finally, the Chain-of-Emotion architecture was implemented into a custom-made conversational game to evaluate against No-Memory and Memory architectures in a user evaluation study (Experiment 3). It was found that the Chain-of-Emotion architecture implementing appraisal prompting led to qualitatively different content generations quantified via the LIWC that outperformed the other conditions on multiple user experience items relating to agent believability and observed emotional intelligence of the agent.

Overall this study provides early evidence for the potential of language model agents to understand and simulate emotions in a game context.

8.5.1 Emotional Intelligence in Language Model Agents

As more and more evidence arises for the potential of language models to simulate cognitive processes of humans [360], I investigated how this could translate to more affect-focused tasks, specifically emotional intelligence tasks. OpenAI's GPT-3.5 performed well overall in situational emotional labelling, providing some evidence for the utility of such models to identify the most likely emotional reaction for a range of situations. These findings therefore add to the body of evidence indicating that language models could be useful to better understand cognitive processes [399]. Importantly, our findings do not only show that LLMs can solve emotion labelling tasks much better than chance level but also that the performance is dependent on the underlying prompting strategy. Adapted from successful chain-of-thought prompting strategies [379], we compared prompts without context (No-Memory) to prompts with previously answered questions included (Memory) and to prompts that first ask the model to appraise the situation and then answer the STEU item (appraisal prompting). This third strategy was built upon findings of modern psychological research that show that cognitive appraisal processes are important factors when it comes to human emotion elicitation and understanding [28], [36]. In a recent study by Li et al. [400], emotionally charged prompts have been shown to improve task performance for various language models. Similarly, appraisal-prompting led to better performance in the emotion labelling task compared to the other two conditions. This finding can be considered from two perspectives: first, it shows that commonly observed psychological processes might be represented in language and therefore in large language models, providing more evidence for the utility of such models to simulate human responses [360]. Second, techniques built upon such observed psychological processes can be used to improve language model performance for specific tasks and might therefore be considered when it comes to building architectures for language model agents. Especially this second point could be of relevance when considering how language model implementations could be integrated in the future to solve problem-specific tasks. Since performance can be increased through prompting strategies facilitating few-shot learning [369], [379] and language models demonstrate representations of a range of psychological constructs [362], building prompts on these cognitive processes is likely to yield benefits for various tasks.

From a psychological perspective, appraisal has long been acknowledged to be a central part of emotion formation, involving both conscious and unconscious cognitive processes [28], [36]. In its basic definition, appraisal relates to an individual’s relationship to a given event in terms of different variables (such as personal significance, goal congruence, etc. [89]). It is not yet clear what specific variables are of importance and how the process of appraisal interacts with other emotion components on a detailed level [28]. That is to say, appraisal cannot yet be universally modelled and therefore implemented within a computational system. We may however assume that information that makes the appraisal process observable and usable is represented in language and therefore also in LLMs [401]. It can therefore be argued that language models could solve some of the practicality problems present in the discipline of affective computing [42]. If LLMs can solve EI tasks through mechanisms mirroring appraisal, we can make use of these models to potentially build affective agents [39] without the need to fully solve the remaining theoretical problems in the field of psychology [28]. The use of language models could therefore be considered a more practical solution to producing useful agents, even if open questions regarding the understanding of human emotion remain.

8.5.2 User Interaction with Chain-of-Emotion Language Model Agents

Implementing appraisal prompting into a Chain-of-Emotion system (see Fig 8.1 for a schematic representation), it was possible to test output contents as measured with the LIWC against other LLM architecture implementations (see Fig. 4 for details on each condition). For the purposes of this study, the implementation was kept as simple as possible and only included text storage (memory system) and an appraisal-prompted emotion generation (Appraisal System) before character dialogue was generated. Within a custom-made role-playing scenario where the agent was used to play out a breakup scenario with their long-term romantic partner, the Chain-of-Emotion architecture demonstrated a higher Authenticity score when prompted with controlled prompts that were kept fixed between all conditions. When tested with players in a custom video game, the Chain-of-Emotion architecture led to a significantly different Tone score of the language, potentially signalling the inclusion of more complex emotional responses as observed in the controlled environment. It is important to note that authenticity was only increased with controlled prompts and tone was only different with non-controlled player-generated prompts, meaning that the differences in text-generated content was highly influenced by the in-game

context. The texts generated for the fixed prompts (see Appendix A.3) yielded potentially more complex emotional responses (for example a mix of melancholy and nostalgia) in the Chain-of-Emotion condition compared to the other conditions.

This pattern was also observable within the user ratings. The Chain-of-Emotion agent was rated significantly more natural and responsive than the agent in other conditions, and additionally more sensitive to the emotions of others. Other items relating to believability and observed emotional intelligence also showed trends of better performances for the Chain-of-Emotion condition. Building such an architecture has therefore quantifiable benefits when it comes to the user experience of artificial agents, which is one of the most important evaluation criteria, especially in the domain of video games [402]. Importantly, no differences in personality aspect ratings (on the classic domains of competency, warmth, capability, and friendliness) were observed. This could be seen as evidence that all implemented language model agents followed the task of role-playing the given character with the provided personality. However, the Chain-of-Emotion architecture outperformed the other architectures in terms of observed emotional intelligence items (quantified via the STEU score) and believability (quantified via user rating). The proposed architecture therefore yielded convincing results on multiple evaluation criteria (qualitative characteristics of content, user-rated believability, user-rated emotional intelligence, in addition to the previously tested emotion understanding) and can therefore be seen as a step towards well-functioning affective language model game agents that could solve some of the problems present in the field [39]. Most importantly, because language model agents have the ability to simulate human-like cognitive tasks [360], a successful game agent architecture does not need to solve fundamental problems in theoretical psychology before creating computational implementations as previously considered [41], [42], [47]. Rather, a language model agent architecture needs to make use of the characteristics of LLMs and implement systems solving more practical concerns, such as memory tasks (both storing and retrieval [374]), or performance-enhancing tasks, such as the proposed appraisal-prompting step. Future research can expand these efforts and test more complex systems and varying game mechanics and user interfaces.

8.5.3 Limitations

Language models do not simulate the behaviour of a human but provide probable language outputs that in some form represent human behaviour. That is to say, models are bound

to their statistical capabilities even in a theoretical, infinitely trained model [403]. This means that while there is no doubt of the potential of language models to solve some tasks with human-like performance [360], other tasks (e.g., truth-telling [403] or casual reasoning [360]) can pose difficulties. As human affect is a very complex field with various competing theoretical perspectives, LLMs cannot be seen as accurate simulation machines of affective human processes. Rather, the provided results show that some psychological processes can be simulated through their representations in language that can be replicated through deep learning techniques. Our application is built specifically upon the process of appraisal [36], which is a well-researched component of human affect. There are however many other processes involved, including physiological, behavioral, and feeling components that are points of contention among theorists [28].

Another limitation stems from the simplicity of our implementation. To test this new kind of architecture for affect simulation, we chose to make all systems as easy as possible. As shown in the study by Park et al. [374], generative video game agents benefit from certain implementations of memory systems that can store and retrieve information with relevancy for the given situation. Since the tested game was rather short and all interactions had relevancy, we did not include a memory-retrieval step, which might be necessary in longer and more complex games. Similarly, the output was limited to the chosen characteristics and context constraints (i.e., the dating scenario) and any additional challenges posed by other contexts/characterisations prevalent in video games (e.g., fantasy settings, historical settings, science fiction settings, etc.) might need further experimentation. LLM architectures are getting more and more involved with techniques such as chaining [404], [405], dynamic context [406], function calling (and therefore more complex decision making), and more. The experiments presented here were designed to present foundational insights about the chain of emotions but fully realized affective agents need to be embedded in more complex systems and further evaluated in their ability to simulate emotional reactions. Furthermore, this study made use of only one LLM (namely OpenAI’s GPT-3.5) as we had no access to other models. As described in some early reports (e.g., [407]), newer models such as GPT-4 likely outperform previous models on various criteria, which likely influences the impact of strategies such as appraisal prompting and Chain-of-Emotion architectures. However, given the domain-specific aims of game character simulation, it can not be assumed that game companies will want to make use of the most powerful language models in every case. Providing strategies for improving language model capabilities will

have value in any case and should inform the process of creating and using appropriate models to solve emotion understanding and simulation tasks in the future.

8.6 Conclusion

This chapter takes early steps towards understanding the role of LLM within affective video game systems. It adds to the body of research showcasing the capabilities of LLMs to solve psychological tasks, namely emotional intelligence items and simulation of believable emotions and how these capabilities could help progress affective game systems. The emotional information represented through the training data in language models seems to hold the necessary information that makes inferring plausible affective states to others possible, corroborating current results showcasing the theory-of-mind abilities of LLMs [361]. The performance is dependent on prompting strategies - utilizing appraisal prompting increases emotional intelligence scores. Based on these results, a Chain-of-Emotion architecture for conversational game scenarios was constructed that implemented an appraisal step to simulate emotional responses and used the resulting information for dialogue generation. This new architecture improved the agent's performance in believably simulating complex emotions (in addition to the previously shown benefits for emotion intelligence). The proposed system can therefore be seen as a successful step towards demonstrating the role of LLMs in the emotional player-game loop, specifically as affective game agents implicitly solving the three main affective computing tasks (emotion measurement, adaptation and elicitation [41]) purely through semantics. It is still necessary to test and refine such systems and implement them within more complex LLM agents and video games. By further developing the efforts presented here, fully realized affective artificial agents could be part of the affective loop and solve some of the problems currently present in affective computing studies.

Chapter 9

General Discussion and Conclusion

This thesis discussed affective games in multiple ways. Following the findings in Chapter 3 that suggested the need for better research integration (specifically modern psychology research into affective computing work), the Emotion Design framework was constructed. It provided both a theoretical model of the affective player-game interaction and a practical guide to develop and evaluate affect-adaptive video games (Chapter 4). The following chapters made an effort to evaluate the framework itself by applying it to the design process of a video game (Chapter 5), to the research of play behaviour (Chapter 6), to the examination of player affect modelling (Chapter 7), and to the development of AI-driven affective agents (Chapter 8). This chapter will summarize the methodological approaches and findings of these empirical efforts and discuss general limitations and avenues for future research.

9.1 Summary of Empirical Work and Contributions

9.1.1 Chapter 3

Affect-adaptive games have gained popularity in recent years, promising benefits for player experience, performance, and health [44], [45], especially in biofeedback-based games [40]. However, our understanding of the exact nature of these benefits is still incomplete due to mixed methodologies and theoretical definitions of "affect-adaptation".

This chapter included a systematic review of research on what kind of theoretical basis affect-adaptive games are being conceptualized, what kind of beneficial effects of playing affect-adaptive have been found, and how studies are being evaluated. Relevant high-quality evaluation studies of the effect of affect-adaptive video games on various outcomes

in regards to their effects, theoretical assumptions, and methodologies were therefore being evaluated. Out of 3,930 papers, 26 studies were included based on preregistered inclusion and exclusion criteria.

The reviewed literature highlights positive effects on various player experiences measures (including enjoyment [204], [206], efficiency [192], [193], and challenge [205]), performance and physiological measures [203], [208], [209], and even learning [212]. However, while many of these results were promising, they were often based on highly differentiating assumptions about emotion conceptualization (especially distinct vs. dimensional emotions), measurement labels (e.g. stress [152] vs. arousal [190]) and measurement techniques (e.g. facial expressions [191] vs. physiology [194] vs. questionnaires [262]). Furthermore, methodological rigour was uneven across studies and 38% ($N = 10$) of observed studies demonstrated medium to high risk of bias (RoB) according to the Cochrane Collaboration's tool [179].

While there have been other systematic reviews in the same field (see for example [40], [44]), none have been found that focus on how specific emotion theoretical approaches are conceptualized in current research, as well as what affective stimuli are being measured and how these measures are being evaluated. The findings of this study have revealed three important observations when it comes to studies investigating affect-adaptive games:

1. A wide variety of outcomes is used to evaluate adaptation effects, mostly with positive associations (especially for outcomes relating to player experience). While we cannot draw universal conclusions given that outcome variables were overall imprecisely defined, affect adaptation seemed to lead to a positive trend in various PX dimensions.
2. Many different emotion theoretical assumptions are being made when evaluating games and many different emotion measurement instruments are being used for the purpose of adaptation. While we cannot know how emotions should be conceptualized, we need to at least validate if our emotion measures reflect our interpretation of the players' affective states. Only a small sample of the investigated studies directly tested if the emotion measure reflected the underlying concept of emotion [194], [195], [210], [215], [262] and even fewer studies additionally validated if the game material used for adaptation would lead to the expected emotional reaction in players [194], [195], [210]. Because of this, it is very difficult to compare different affect-adaptation mechanisms and therefore almost impossible to draw broader conclusions.

3. Finally, while all investigated studies were high-quality studies defined by a randomized-control design, a third of the investigated studies showed at least some form of risk of bias [179]. This finding indicated that the research field of affect adaptation may need more studies adhering to methodological research standards for controlled trials.

Overall, this study contributes quantification of theoretical and methodological barriers present in affect-adaptive game studies. While demonstrated effects seem promising, it's not possible yet to draw meaningful conclusions that are useful for actual game development based on the studies investigated here. One way to potentially address this is the development of clear standards and theory clarifications that help make affect-adaptive games robust and comparable.

9.1.2 Chapter 4

The Emotion Design framework was constructed to specifically address the barriers identified in Chapter 3. It consists of two parts: (a) An interactional model of the affective data flow between players and adaptive game systems; and (b) a practical framework with specific steps to follow to construct robust and validated emotion adaptation.

The interactional model was based on the affective feedback loop [27], [47] and expands the concepts by explicitly including psycho-physiological affective components on the player side of the game-player interaction. These components include cognitive aspects of emotion elicitation (specifically appraisal [36]), behavioural, physiological, and subjective feeling components which represent the main component of interest when designing and developing emotionally evocative games. It also provides new labels for affective information presented by the game to elicit player emotions (the domain of emotional output), as well as for affective information measured by the player to adapt the game (the domain of emotional input). It provides therefore more structure and clarity to emotion theoretical concepts and how they relate to each other within human-computer interactions. One important aspect of the model is the notion that emotion elicitation and expression are highly dependent on individual differences [90], [111] and context [105], [106]. It is therefore necessary to validate game material (potential emotional output) and player measures (potential emotional input) by testing associations with subjective player experience for the target players within a specified game context.

The methodological framework that accompanies the interactional model guides crucial standardized steps towards making solid, comparable affective games that explicitly test

any theoretical affective assumptions to improve the overall reliability and validity of the development process of such games. This process functions within two domains:

- **First Domain: Design Emotional Output.** Emotional output relates to stimuli potentially eliciting the target emotion. The standardized process must include explicit testing of the game stimuli for the potential to truly elicit the target emotion within the target audience.
- **Second Domain: Integrate Emotional Input.** Emotional input relates to player data that might indicate the target emotion. The standardized process must include explicit testing of the user data and how it predicts the target emotion for the target audience.

The framework was created to help construct affect-adaptive games that are comparable and not reliant on potentially outdated or oversimplified theoretical assumptions. It is argued that it can help designers and researchers conceptualize affect adaptation for all kinds of video games and plan the development process in a standardized manner. Importantly, the framework does not claim to solve all problems present in the field (especially given the fundamental theoretical uncertainty in the field of emotion research [28]). However, it provides concepts, labels, and tools for developers and researchers to gather and evaluate all necessary information within the affective player-game interaction and construct affect-adaptive systems according to our current knowledge in the domains of psychology, design, and computer science.

9.1.3 Chapter 5

The Emotion Design framework is based on wide-ranging literature research and the empirical data found in Chapter 3. However, the value of the introduced concepts and methodology can only be assessed if it is thoroughly empirically tested. Chapter 5 was constructed with two experiments to evaluate the process model within the development cycle of an affect-adaptive video game.

In Experiment 1, 161 participants played through a newly developed 2D action arcade game. Their data was used to identify associations between game input, presented material, and affective valence. In the context of the game, three potential emotional output variables have been taken into consideration: enemy attack speed, musical tempo, and synchronization between attacks and music. The results demonstrated that attack speed

had a significant impact on the subjective target emotion (valence) and was deemed an appropriate emotional output for adaptation. A prediction model implementing the number of blocks and number of deaths as predictors for subjective valence has been identified as the best-fitting model and therefore as the best representation of emotional input for this game. The main contribution of this study is to showcase possible use cases for the methodological framework. Specifically, this study assessed affective associations between 3 possible emotional output variables (enemy speed, musical tempo, and music synchronization), but only found one significant effect (enemy speed predicting subjective valence). Even though the other candidate variables were based on foundational research [270]–[272], within the context of the game participants were only meaningfully influenced by enemy speed when quantifying valence. Similarly, 8 in-game variables were considered as potential emotional input (indications for the player’s affective state), but the best-performing regression model only included 2 (number of blocks and number of deaths) to best predict subjective valence. This process therefore demonstrates the importance of a user research step given a specific game and player base.

In Experiment 2, 158 participants were randomly allocated to play either one of two new adaptive versions of the game or a non-adaptive control version. The player-based adaptation used the previously identified prediction model (which used the number of blocks and number of deaths as relevant data sources) to inform when and how attack speed would adapt to increase subjective valence. The level-based adaptation changed the attack speed purely based on the currently played level. A significant increase in self-reported valence was found for the player-adaptive game version compared to both other conditions. This study therefore provides evidence that this process can be utilized to create effective affect-adaptive video games. It further demonstrates that such an adaptation is not dependent on physiological data (which is the standard for affect adaptation [39], [40]), but can be achieved by in-game behavioural data of players.

9.1.4 Chapter 6

"Emotional output" in the context of the Emotion Design framework relates to game material with the potential to elicit a target emotion in the player. From a broader perspective, it bridges the complex nature of modern video games to the even more complex psychology of human emotions. This construct explains a fundamental part of the player-game interaction and helps us understand real-life observations that involve play behaviours. One

such observation is the demand for certain types of video games during the COVID-19 pandemic.

The COVID-19 pandemic has been a major source of stress for a majority of people [289], [290] with potentially negative long-term effects on mental health and well-being [294], [297]. There have been recent reports of potential positive effects on stress relief in video games, especially games that relate to the target emotion of “relaxation” in some form [305], [306]. Utilizing the Steam relaxation tag to group individual games into "relaxing" and "non-relaxing" games and matching sets by player counts pre-pandemic, a quasi-experiment was conducted investigating the average daily player peak (ADPP) for the COVID-period against the pre-COVID period for a matched test set of 143 video games. A medium-sized increase of ADPPs over all types of games but no difference between relaxing games and non-relaxing games.

This study provides insights about player behaviour during a highly stressful time. The overall rise in video game play fits prior research showing increased screen time during the pandemic [309], [310], but it also uncovered that this was not more the case for "relaxing" games. In fact, different to many other studies focusing on "relaxation games" to evidence the stress-relieving aspect of video games (e.g. [339], [340]), this study showed that relaxation is not necessarily a characteristic of a game, but rather a characteristic emerging from game-player interactions. This study provides further evidence of the usefulness of the emotion design framework as it conceptualizes this interaction and is useful in providing possible explanations for such effects. Emotional output as game material evoking a target emotion is dependent on the audience - on context and individual differences specifically. Furthermore, this study showcases how large-scale analyses of available gameplay data can be utilized to examine human behaviour in highly stressful times. How humans interact with games can indicate a deeper truth about the nature of behaviour and emotions [37], [408].

9.1.5 Chapter 7

The counterpart of emotional output is "emotional input" and relates to player data that provides some indication of the experienced target emotion within the player. There are behavioural, physiological, motoric, and even semantic aspects of human emotion that can be parsed into a game system to indicate affective states. But similar to emotional output, emotional input cannot be put into a universal schema that takes in human data and out-

puts an emotion for every person in every situation alike. To illustrate this, this chapter expanded on an experiment by Sullivan et al. [346], where 64 participants were allocated into a sleep-deprivation and sleep-rest group and then played through a 10-minute custom VR horror experience while their physiology and positional data were measured. A mixed linear regression model was used to predict the subjective valence of each participant throughout the experience. It was found that these valence measures were predicted differently between groups and that positional measures interacted with the group in how well they predicted valence. It was furthermore found that physiological measures did not contribute meaningfully to valence prediction, further indicating the limitations of physiology measures in predicting affective states universally [32]. This study therefore provides evidence that the usefulness of physiology and position measures are dependent on many factors and cannot be assumed "objective universal measures" of affective states. While physiology did not improve the prediction model at all, position measures did significantly predict valence. They also contributed to valence prediction in interaction with the group condition, demonstrating that the statistical relationship between body movement and subjective experience in this VR horror world was at least partly moderated by the amount of sleep the participants had. Again, these effects can be explained within the emotional feedback loop that indicates the necessity to validate the influences of individual differences and context to specific emotion measures. Emotional input as player data indicating a target emotion is dependent on the audience and the situation the audience is in. Broadly speaking, these findings add to the body of research showcasing the complexity of emotion elicitation and measurement [28], [32], [97], [254] that make one-to-one mappings between specific affective states and measurement outputs questionable. Emotions are embedded into a complex system with many components that are influenced by our surroundings and our internal states [36], [38]. However, it is not relevant to understand everything to build a useful statistical model of a game-player interaction. As this study shows, we cannot make universal assumptions, but we can try to find useful emotional input sources to model affective player states. Again, the emotional feedback loop can help us understand this process and relate our observations to known components of the emotional player-game relationship.

9.1.6 Chapter 8

Generally speaking, the emotional player-game relationship is an important concept for all kinds of games given that fundamentally games are about creating experiences [21]. With more technological progress, new types of affective games become possible that still are represented through an emotional feedback loop. Making use of evolving technologies is an important part of game development and understanding the interactions between players and games when it comes to affective information can help further applications of new technologies and therefore evolve what affective games are capable of. Large language models for example have shown a potential to understand and simulate human-like behaviour purely from semantic information. Given that semantic information plays a big role in how we conceptualize and view emotions [38], [409] and how cognitive processes play a role in emotion elicitation and expression [29], [36] large language models hold potential to power successful affective systems that interact with users on a semantic basis.

To test the theory that language model agents would also be able to simulate an understanding of appropriate emotional reactions, OpenAI's GPT-3.5-turbo model was used to solve emotional intelligence tasks through the Situational Test of Emotional Understanding (STEU; [386]). Results showed that the language models were able to correctly solve 57% of the tasks in a baseline condition, but could be improved to solve 83% of the tasks by adapting the prompt for the models to include instructions to appraise the situation first before indicating the appropriate emotional reaction, mirroring cognitive processes in human connected to appraisal [36]. This chapter contributes very early evidence that LLMs have the ability to not just emulate purely cognitive processes [360], but also affective processes. It furthers our understanding of technological capabilities, but also of how mental constructs are represented in language in a way that makes simulation possible through complex machine learning techniques. Furthermore, considering the emotional feedback loop, this study demonstrates that designing an LLM-based emotional understanding system can be enhanced by mimicking the affective information flow in the player component.

In a second and third study, a new LLM game architecture was created that implemented an appraisal system within a game agent to make semantic emotion simulation possible. This Chain-Of-Emotion architecture which builds upon design patterns like chain-of-thought [379] significantly outperformed control architectures on a range of user ratings (notably believability and sensitivity to emotions of others) and content analysis scores

(authenticity). This architecture can be considered an early example that purely semantic LLM-based game systems can effectively adapt to human input in an affective game scenario and create believable emotion simulations. In this example, both emotional input and emotional output were represented as text in a conversational video game, which led to an emotional player-game interaction mimicking real human-to-human conversations. It therefore contributes to our knowledge of how LLM could be used to build effective emotional systems and also uses prior work represented in the emotional feedback loop to make improvements possible. With new technologies, new types of affective game systems will be possible, and if properly integrated into our knowledge about the human psyche and player-game interactions, these systems show improvement.

9.2 Progressing Affective Systems

As stated in the Introduction of this thesis, the main objective of this thesis was to further our understanding of affective interactivity in video games by providing a framework that addressed theoretical and methodological barriers in the field. The Emotion Design framework was created to specifically target such common barriers, including theoretical uncertainties in fundamental psychological research, conceptual fuzzyness when it comes to relevant affective game components, unclear process steps when developing affective games, and a lack of standards and criteria in affective game evaluations.

By the nature of the framework, it is meant to be applicable in various contexts where these barriers pose practical problems. The following section will outline some of these usecases explicitly.

9.2.1 Emotion Design for Game Designers

In game design practice, the concepts introduced in this thesis can hold immediate potential benefit by structuring the necessary parts of the emotional game-player interaction that could be relevant for a certain design. For example, through the concept of emotional output, designers gain the tools to describe and explore specific game stimuli and their ability to elicit certain player emotions. Emotional output is defined as any material that has the potential to elicit a target emotion and the elicitation process is outlined as a multi-componential affective process involving appraisal, individual differences, and contextual differences. Therefore, a specific feature that needs to elicit a certain emotion can

be described and tested with these factors in mind. In Chapter 5, The Flow Experience made use of this by specifically testing the relationship between enemy attack speed and enjoyment measured as subjective valence for individual users. Another game might specifically test the relationship between a puzzle layout and frustration or a cutscene event and fear. In other words, what in practice might often be addressed implicitly [129], [235] can be more clearly defined and explicitly tested through the framework.

Similarly, emotional input is defined as any data that might be used to infer player emotions. A game design that relies on optimizing a specific experience can through the framework directly identify and test candidate data sources. The Flow Experience made use of this by modelling subjective valence with in-game player data as predictors, which resulted in an adaptive version of the game that successfully increased the enjoyment (see Chapter 5). Another game might want to identify useful data sources to measure sadness in an attempt to optimize for a certain narrative experience or identify relaxation to provide gamified emotion regulation trainings. To achieve this, the framework provides relevant constructs, explores key literature, and describes possible measurement instruments to enable game designers to approach their specific needs in a way that avoids common pitfalls (see, for example, discussions of the applied affective games in Chapter 3).

9.2.2 Emotion Design for Researchers

The origin of the theory-practice gap discussed in Chapter 3 can be traced back to the disagreement and unknowns present in fundamental psychological research. While countless findings have shaped our understanding of emotions in the last years, there is still much to learn and even basic theoretical disagreements are not yet resolved [28]. Applying emotion theories in practical applications, such as game applications, could help answer some of the open questions present in the field.

The Emotion Design framework was designed to bridge our current scientific knowledge with real design practice. Potential usecases therefore include the design and development of affective games discussed in the previous section, but also applied game research. Chapter 6 examined player behaviour during the COVID-19 pandemic and therefore more generally emotion regulation in highly stressful times. Through the lense of the Emotion Design framework, patterns found in the large body of player data helped identify new knowledge gaps and propose possible explanations of how video game characteristics relate to individual stress management strategies. Chapter 7 on the other hand researched

emotion modeling techniques in a complex VR environment and how circumstantial player characteristics could impact measurements.

The player model and emotional information domains described in the framework helped to put these two studies into context and guided part of the methodologies (particularly data acquisition and analysis of Chapter 7). It can be used to frame theoretical assumptions, explain findings, and guide the process of building applied game scenarios for research purposes.

9.2.3 Emotion Design for Affective System Engineers

With the rise of intelligent systems, a new interest has emerged in affective systems, not only for video games, but for computers in general. Traditionally, the development of affective systems was limited by both theoretical certainty and technological advancements [42], [142]. The Emotion Design framework identifies and acknowledges these gaps and provides consolidation work to guide affective system developments given the current state-of-the-art.

Chapter 8 provides an example of how a new large language model affective system could be conceived, implemented into an interactional game, and evaluated for relevant factors. More specifically, the chain-of-emotion architecture builds upon the affective system model of the Emotion Design framework by simulating appraisal processes (which are conceptually intergrated in the affective player model) through a large language model architecture. This led to a functioning, believable affective agent that engages on an emotional level with the player. Emotional information (emotional input and emotional output) was conceptualized semantically and supported both the development of the system itself and the evaluation criteria.

Again, the consolidation efforts of the framework bridge knowledge from emotion research and design practice and therefore provide some fundamental concepts to make progress in the field of affective systems easier. Building upon these efforts, new affective systems could take these concepts into consideration of more robust systems (e.g., simulating affective individual or cultural differences in emotion elicitation). Concepts of emotional output and input can guide evaluation criteria (e.g., does the system emulate the desired affective reaction? Does the system measure the correct emotional reaction from the user?). The framework itself is deliberately fluid to allow for both theoretical and technological progress.

9.3 Limitations and Future Directions

The Emotion Design framework was constructed out of the findings of affective game evaluation studies showing inconsistent theoretical and methodological approaches. It has been applied to multiple different types of games and interactive scenarios to present use cases and validate the presented process. However, this work comes with several limitations that need to be acknowledged to outline potential paths of future research.

9.3.1 The Emotion Design Framework

The emotional feedback loop addresses some of the concerns identified in Chapter 3 by integrating modern psychological and HCI theories into one robust, interactional model of the affective player-game interaction. In doing so, it makes use of contemporary knowledge of emotions to construct a player component and contemporary knowledge of affective systems to create a game component and connects these via two domains of affective information.

One major barrier to applications making use of psychological theories is that these theories often involve simplified, abstract representations of complex processes. Emotions in particular have been extensively studied but are still not well understood [28], [97]. This means that fundamental aspects of emotions could be revealed in the future that completely change our definition of how a player component should function. The framework integrates emotion components that have been agreed upon by most theorists [29] and focuses on giving developers robust tools to make affective user research themselves. Its robustness, therefore, comes from its limited use of hard assumptions. Emotions can be conceptualized as dimensional representations or as distinct categorical states while integrated into the framework. The aim was to make the framework useful despite our limited knowledge, which naturally leads to limited precision in terms of theoretical definitions. The framework cannot answer questions even the most current science is unsure about. It cannot be used to explain complex psychological processes in detail. Instead, it is an open model and a toolbox that can be iterated with further progression of knowledge. This is not only true for the player component, but also for the game component. The feedback loop defines three modules of the game component: A detection module, an adaptation module, and an elicitation module. These exist only to define basic functionalities of affective systems, but do not provide a proper architecture to build them, since they need to be open to further technological progress [47].

The framework is therefore created with an emphasis on practicality, although the process presented in this thesis does come with a cost. Both domains of affective information introduced in the framework (emotional output and emotional input) are necessary for adaptivity, but contemporary research shows that there are few assumptions regarding these domains that hold true across populations and contexts. As a consequence, user tests become an integral part of fundamental design decisions. Depending on the target emotion, there are few validated measurement instruments, so exploratory fundamental research is necessary to even find ways to detect the target emotion, which might not even be practical in a production environment. Depending on the difference between players, some games might need to implement very complex player models that are costly to develop and run and might have limited benefits [410]. Still, as user research is an important part of many game projects already [241], the additional resources necessary to create a more emotionally adapted game compared to potential gains [47] really depends on the project and the aims. While cost could exponentially grow when adapting every single game element to every possible player and situation, there are many examples of the benefits of some kind of emotion adaptivity, ranging from broad game aspects [44] to complex procedurally generated game worlds [158]. Adaptivity can bring a game closer to the intended experience but can become very costly for high levels of personalization. Since humans are not purely reactive, but active participants in constructing their experiences [38], it can be argued that the main potential of the process lies not within completely personalized experiences for every player in every situation, but for specific gaps between intended experience and observed experience. As of now, there is more work necessary to gauge how costs and benefits balance for different types of projects and levels of adaptation and how the framework can evolve the more knowledge we achieve in the fields of psychology, design, and affective computing.

This specifically means that the value of the framework for practitioners can only currently be assessed by the examples given in this thesis. In the previous section, I discussed potential usecases but a necessary limitation of this thesis is the lack of concrete guidelines. There are countless design practice approaches, potential research projects, and engineering efforts that could benefit from concepts introduced in the framework on a theoretical level. A robust integration is however still in need of more work: How should the practical guide in the framework applied to large scale game projects to optimise cost and benefit? How can academic researchers make simple use of a robust affective game

to research fundamental questions about the human psyche? Are evaluation strategies outlined in this thesis sufficient for assessing new and progressive affective technology? The most immediate research gaps lie between the introduced concepts here and specific, actionable sets of guides. As described, game designers can use the framework to gain crucial understanding of the emotional potential of their games and the affective reactions of the players. Integration of this knowledge in design practice needs to be examined in terms of scale, practicality, cost, generalisability, and proposed benefits in order to create in order to concretely guide this process in the future.

Whit that in mind, this thesis provides various data-driven examples of how the proposed workflow can benefit game development and research in various ways. As it stands, these examples are not scalable to all aspects of industry or research practice but provide glimpses of the potential of the provided consolidation work that should be expanded into various concrete practice examples to maximise the proposed benefits. The limits of the framework in terms of applications are not clear, as is the potential for iterating and improving the provided concepts. However, the studies collected here show a clear potential of the framework to guide progressive work in affective system development and research on a broad level. It that sense more targeted application guidelines for specific use cases are likely to enhance the potential gains of the Emotion Design framework.

9.3.2 Understanding Players

Multiple studies have been conducted that make use of the Emotion Design framework to build or analyse emotional interactions between players and games. On a general level, these studies show the potential of the framework to (a) build successful affective games and (b) understand observations relating to player affect. However, it is important to note that these studies do not allow for definite conclusions about the exact nature of the observed results. In fact, it can be argued that there are too many unknowns in our understanding of emotions to provide precise explanations of many of the findings. A deeper exploration of the relationship between emotion components is currently a central topic in psychological research, which includes some new discussions about how many distinct emotional states exist and how they can ideally be labelled and differentiated [35]. Progress in this field will lead to a better classification and description of emotions, as well as an improved way to design for specific emotion elicitation. A more elaborate and accurate model integrated with the framework would aid the development of a complete emotional engine

in games that could simulate emotional states (for example in NPCs) or reflect the real-time affective reactions of players [39], [142], [266]. Presently, the psychological basis for many observations made in this thesis is not clear, for example, why the number of blocks and the number of deaths best predicted valence in Chapter 5 or why relaxation games did not differ from other games in player peak growth during the pandemic in Chapter 6. The Emotion Design framework acknowledges these observations and implements them within a useful methodological process, but it cannot provide definitive answers. The most pressing conclusion the framework helps to find is that developers and researchers need to take actual player experience into account and must explicitly test their assumptions about emotional input and emotional output. It is currently necessary to test associations between game characteristics, player input, and emotional experience with consideration to individual differences and context. It may be possible that certain game characteristics have universal affective effects or that certain measurement techniques help recognize universal affective patterns. However, the presented evidence strongly points towards the notion that we cannot blindly assume universal bindings between video game material, player input, and target emotion.

This does not mean that the presented findings cannot be used to infer information about broader human behaviour. It is an explicit aim of this work to help shape our knowledge of human experience in interaction with affective game systems. How we interact with games can lead to more knowledge about psychology in general and as shown in Chapter 6 and Chapter 7, the framework can help construct research to better understand human emotions in game interactions. Evidencing the importance of user research, individual differences and context is however only the first step. More research is needed to fully understand the intricacies of emotion elicitation and expression within the player-game loop. Every study provided here leads to interesting further questions that need more research to address: What are the mechanisms that lead to the same stimuli leading to different emotions in players? What are the mechanisms that lead to player state (in the context of this thesis specifically sleep deprivation) leading to different mappings between body position and subjective experience? This thesis contributes many interesting observations of affective player-game interactions, but cannot explain these in detail. It should therefore be seen as a first step towards better (and more standardized and comparable) research in the field. In this sense, this thesis demonstrates that games can provide important data to increase our understanding of emotions. As the process of designing for

emotions gets better understood, the relationship between appraisal, emotion components, individual differences, contextual differences, and how emotions evolve over time will have data-driven evidence for specific (and potentially very complex) target emotions. We only begin to understand the satisfaction of negative emotions or the gradual differences between states and games offer a great interactive environment to explore these emotions in an experimental setting. Games also provide a unique opportunity to process large sets of emotional data in a bottom-up approach, meaning emotion models can be created by investigating patterns in raw data [411], [412]. A mixed-methods approach of theory-driven and data-driven emotion models investigated through games could provide much-needed empirical evidence that brings us closer to a comprehensive understanding of the nature of emotions.

9.3.3 The Future of Affective Games

One explicit aim of this thesis was to advance the affective design process for video games as it could lead to intense emotional experiences and affect-aware games. Future technologies will open up implementation possibilities, such as affect-based procedural content generation (PCG), i.e. the automatic generation of game elements through machine learning algorithms that are based on the momentary emotional profiles of the players. Work already moves in a direction to build stimuli automatically based on models of player experiences and this approach will offer many possibilities for fully or partially automated game design practices that are emotion-adaptive [158], [231]. With more and better measurement techniques, it will also be viable to create new possibilities of game interactions. More and more game interfaces are being developed, like portable EEG systems or face recognition software from webcams [47] that will only improve the ways people interact with games. Recent empirical studies already showcase the potential for affective interface design in regards to user experience [157], and the normalization of sensor integrations, like eye gaze in VR or EDA sensors in controllers, will open up many more possibilities to explore [257].

Chapter 8 demonstrates how a completely new technology can be used within the emotional feedback loop to create new kinds of affective systems. It provides a useful architecture based on the framework to create affective LLM agents. Because many of these new avenues are fairly young and rapidly evolving, we cannot yet provide a deep exploration of how to best construct affective systems. Chapter 8 in particular provides a first, research-

driven step towards further investigations of how an LLM-driven affective engine might look like in the future. The potential of affective video games in multiple different forms can be observed from each empirical chapter of this thesis, but full implementation details remain a task for future research. The main aim of this thesis is to provide theoretical models and technical process illustrations to avoid unfounded theoretical assumptions. The presented work can be seen as robust to theoretical uncertainties, but it does not account for the full spectrum of human emotional responses or technological possibilities.

9.4 Conclusion

This thesis examined affective games, specifically the question of how they function theoretically and how they can be practically implemented. Overall, this thesis provides a theoretical and practical framework for affective games that can be used to examine the emotional player-game relationship and create new kinds of emotional video game experiences for players.

One additional aim of this thesis is to push for better interdisciplinary exchange, which ideally includes theory integration, study collaboration, and data sharing (see [266] for more details on how affective computing research tries to achieve this). As illustrated in this thesis, emotion design emerges through the game-player interaction based on findings in psychology, design, and HCI. Games research has the potential to deeply explore the relationships between these disciplines and connect knowledge that benefits research and industry. As data is desperately needed to fill gaps in all aforementioned areas, games could be used to gather data from specific sensors, objective behaviour, or in-game inputs. There is a massive amount of under-utilized data in games that could help to provide evidence for many more research questions [242], which exceeds the presented areas to other behavioural, social, or even medical disciplines. Emotion research, design, and affective computing are all very complex fields that constantly evolve, making research integration progressively difficult. Still, interdisciplinary research has a crucial role in furthering our understanding of the world. In this sense, this thesis was written with the hope of creating a more unified expression of theoretical work, which could serve as a foundation for achieving closer knowledge exchange.

Appendix A

Supplementary Material

A.1 Chapter 3: The Theory-Practice Gap

Table A.1: Complete list of studies included in the review with summaries of aims, methods, and conclusions.

Authors	Aims/Objectives	Methods	Results and Conclusions
[187]	To develop and evaluate a game balancing system based on facial expression recognition with the aim to enhance player experience.	Two groups of 30 participants (68% male) played either a 2D platformer or 3D shooter in two conditions: Affect-adaptive vs. non-adaptive. Participants answered a subsequent player experience questionnaire [180].	The adaptive versions of both genres showed significant improvements in experience domains such as immersion, flow, challenge, and positive affect. No differences were found negative affect and only for the 2D platformer for competence.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[152]	To prove practicality of ubiquitous biofeedback serious games by developing and evaluating a physiology-based stress management game.	Exp 1: 15 participants (60% male; mean age 33.47 years) played a biofeedback game with stressful/relaxing tasks to test if physiological stress corresponded to game adaptation. Exp 2: 12 participants (58% male; mean age 33.92) played an adaptive vs. non-adaptive version for 5 days each and answered a post-treatment questionnaire.	The game demonstrated a good reflection of physiological stress as presented in experiment 1. Experiment 2 showed that the adaptive version provided a better mental stress reduction over five days. Limitations are discussed in terms of generalizability.
[188]	To develop a mental state-adaptive FPS and evaluate the adaptation against performance-based adaptation in terms of enjoyment and scoring.	21 participants (76% male); age range 19-27; efforts made to validate affective measurement and emotion elicitation through the game before evaluation; participants played affect- and performance-adapted game with HR and EEG sensors and answered adapted GEQ [189] questions (5 items).	Performance-based adaptation resulted in significantly higher flow-experience scores and significantly better performance compared to affect-adaptation. Discussion states small sample size and lack of generalizability as possible reasons, as well as the limited number of predicted affective states.
[190]	To apply emotion-based difficulty adjustment based on facial recognition to a horror game to improve player satisfaction.	31 participants (unspecified demographics) played two versions of a survival horror game: One with difficulty adjustment based on facial expressions and one without. Evaluation was based on number of observed positive and low-arousal emotions.	The adaptive game provided significantly less observed positive valence-emotions and low-arousal emotions, which is argued to show a successful fear experience. Descriptive and qualitative data was provided to show good player satisfaction for the adaptive game.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[191]	To develop an online game difficulty personality system based on Facial Expression Analysis (FEA) and evaluate it within a popular platformer.	Exp 1: 38 participants (47% male); mean age 35.1; participants played through three versions of an Arcade game with FEA sensors to evaluate prediction of perceived difficulty; Exp 2: 10 (without head pose analysis) and 25 (with head pose analysis) (80% male); participants played a static and personalized version of a 2D platformer and rated their preference.	Perceived difficulty was measurable through FEA, which provided the possibility to create an heuristic online personalization system that was preferred by players when used in a 2D platformer, compared to a static game version. Similar results were found for a modelling approach that includes head pose analysis.
[192]	To clarify how affect-based game adaptation can improve implicit recognition of playing styles and performance within a 3D puzzle game.	30 participants (60% male); mean age 31; participants played a 3D puzzle game with and without affective-adaptation controls in a randomized order and answered a post-game questionnaire indicating playing styles and adaptation enjoyment.	Recognition of playing styles yielded a good accuracy within a game combining affective and performance adaptation. The adaptive version of the game showed higher performance and good enjoyment ratings. Limitations are discussed in terms of generalizability.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[193]	To propose and test a linear regression-based model to recognize player styles and test it within an affect-adaptive game.	Exp 1: 34 participants (53% male); mean age 26.85; participants played through an adaptive and non-adaptive VR puzzle game and answered playing style questionnaires [413], [414]. Exp 2: 30 participants (60% male); mean age 31.87; same setup to validate playing-style recognition.	Playing style recognition through affect-related and gameplay data was achieved with an accuracy between 73% and 84% and adaptation based on affective data led to improvement in effectiveness, efficiency, and difficulty of a puzzle game. Combination of affective adaptation and playing style-adaptation is recommended.
[194]	To compare five difficulty adjustment methods in a video game, including manual, random, performance-based, personality-performance-based and physiology-personality-performance-based.	50 participants (74% male; mean age 25.1) played one of five game versions which adapts difficulty: (a) manually, (b) randomly, (c) performance-based, (d) personality-performance-based, (e) physiology-personality-performance-based. Experience was measured through Intrinsic Motivation Inventory [181] and Flow Experience Measure [182].	Physiology-based affective adaptation did not lead to an improvement in game experience, compared to any other group. Physiology-based adaptation may show promising results in validation studies but do not guarantee user experience-improvements, even if all affective relationships are tested in a preceding open loop study.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
Ewing2016	Evaluation Loop validate a psychophysiological model between a player and a game and apply it to an affect-adaptive game.	Exp 1: 20 participants (45% male; age range 19-36) played Tetris with EEG sensors equipped, followed by subjective questionnaires; Exp 2: 10 participants (40% male) played 3 affect-model adaptive game versions and a manual-adaptive version for difficulty adjustment and answered affective and player experience questionnaires [183], [196]	The presented 2-step process to associate physiological data to psychological construct resulted in valid predictions for cognitive demand and effort using EEG measures. The evaluated adapted game showed no improvement in most of the used experience measures. Results are discussed in their utility of a conceptual process model to develop theory-based affective games.
[197]	To propose an approach of emotion-based difficulty adjustment using self-report measures and evaluate it empirically.	66 participants (73% male; mean age 30) played a 2D platformer with emotion-adaptive difficulty, increasing difficulty, and fixed difficulty. Differences are reported in terms of the Intrinsic Motivation Inventory [181] and the Player Experience Inventory [184].	The emotion-adaptive game shows increased player experience ratings compared to both control groups. Additionally, in-game dialogue-based subjective emotion measures showed a high accuracy. Limitations are discussed in terms of more possible comparison groups (such as performance-based).

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[199]	To implement and evaluate an AI experience manager to keep players on a predefined emotion trajectory within a narrative video game.	Exp 1: 294 participants (50% male; mean age 19) played either a game managed by the PACE AI experience manager or by a random model and rated their experience scores. Exp 2: 39 participants (41% male; mean age 20); same setup as Exp 1, but with a preceding calibrating task [185].	Experiment 1 showed no statistical difference between groups, possibly based on a missing calibration as a form of reference. As a consequence, Experiment 2 introduced a calibration task, but again there were no significant differences observed, leading to inconclusive results.
[200]	To test if gestures can be used to recognize emotional states and adapt music to these states using a VR game.	22 participants (67% male); mean age 29.09; participants; participants played either an adaptive (gesture-based fear recognition to change music) or non-adaptive VR game and rated presence via the SUS [201].	Head gesture was found sufficient to detect fear, but no other emotional state. Fear-adaptive music in a VR world was shown to increase perceived presence of players compared to a control game. Limitations are reported in terms of the system's responsiveness.
[202]	To assess the usefulness of physiological data to increase rapport with NPCs in an action RPG.	16 participants (63% male; age range 18-34) were assigned to either a adaptive or non-adaptive game version utilizing EMG and EDA data to change NPC behaviour. Rapport was measured with 3 items on a 9-point likert scale.	Evaluation showed no difference in rapport ratings between adaptive and control group, but qualitative questions indicate enjoyment of the adaptation.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[203]	To propose and evaluate an educational game with affective induction through a fuzzy system analyzing performance and emotional states.	40 participants (unspecified demographics) played both an educational game with linear and with affective difficulty adjustment based on voice recordings. In-game performance and emotional reactions are used as evaluation.	A previously tested emotion classified showed medium to high accuracy for valence and arousal dimensions. The adaptive game showed significant improvements in experience of pleasant-high states and reduction in unpleasant-low states. Adapting both difficulty and aesthetics was considered a promising approach.
[204]	To design and implement an affect-based difficulty adjustment system based on anxiety measures and evaluate its effect.	Exp 1: 15 participants (47% male; age range 18-34) played six sessions of Pacman over two months while physiological data (HR, EMG, EDA) and subjective reports of anxiety were assessed to create an emotion model. Exp 2: 9 participants (unspecified demographics) played both a performance-based and anxiety-based adapted game and answered questions about their anxiety, enjoyment, challenge, and perceived performance.	Anxiety was accurately predicted (88%) with the created emotion model (Regression Tree) through a combination of physiological measures. Significant improvements for enjoyment, challenge, and perceived performance was reported for the affect-adaptive game compared to the performance-adaptive, with no significant difference for reported anxiety.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[205]	To test if facial expression recognition can be used to dynamically balance a game and enhance the experience.	32 participants (unspecified demographics) played both a facial expression-adaptive and non-adaptive game and answered the Immersive Experience Questionnaire (20 items; [183])	After an initial survey, a simple Hack and Slash game was designed with dynamic balancing based on facial expression recognition. The adaptive version showed improvements for the challenge and player experience domain in the follow-up questionnaire.
[206]	To create and evaluate an affective game engine to test how player abilities, enemy design, and environment influences performance and effect.	16 participants (34% male; age range 18-32) played through four game conditions (control, player adapted, NPC adapted, environment adapted) with EDA-based adaptation. Dependent measures included skin conductance response, game performance, and player experience [16], [181].	Results show that the adapted versions of the game were more physiologically arousing, indicating successful arousal-adaptation. Results on player experience scales reveal no effect of adaptation. NPC-based adaptation was reported to be especially ineffective as enjoyment-reduction was observed. Limitations are discussed in terms of generalizability.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[207]	To develop and test a procedural horror game that adapts to affective physiological states.	24 participants (67% male; mean age 22.5) tested three versions of a horror game with physiological measures (EDA, EMG, HR, HRV): A symbiotic adaptive version (in-game character mirrors player affects), a affective difficulty adjustment, and a non-adaptive version. Participants then answered the Game Experience Questionnaire [180].	The adaptive game versions showed improved ratings on the domains of immersion, tension, positive affect, and negative affect compared to the non-adaptive game. It was also reported that game adaptation was successful in shifting player experiences. Further analyses, including qualitative data, provides evidence for the interindividual differences in emotional experiences and elicitation effects of emotion-adaptive materials.
[208]	To develop and evaluate an adaptive biofeedback game that teaches relaxation skills by monitoring players' breathing rates.	9 participants (78% male; age range 22-33) performed a Stroop color test, and then played either a biofeedback relaxation game, performed deep breathing, or played a traditional game. Physiological data was assessed through HRV and EDA during a follow-up stress-inducing task.	The adaptive game was reported to show good skill transfer in terms of relaxation training and showed significant improvement in terms of physiological arousal compared to the other groups. The main benefit of the adaptive game is reported to be the ability to create stressful situations while training relaxation skills.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[209]	To present an evaluate an adaptive biofeedback game for teaching self-regulation of stress.	25 participants (60% male; age range 19-33) tested an emotion-adaptive game using three modalities (EDA, HRV, breathing rate) against a deep breathing treatment and a non-adaptive game after a baseline breathing phase. Performance and physiological data was assessed in a follow-up stress-inducing task.	There were mixed results reported, indicating positive but non-significant improvement in breathing rate and performance for the adapted game versions versus the control, and significant physiological arousal improvement for the breathing rate-adaptive game versus the control game. Results are discussed in terms of the potential of games to manipulate arousal-inducing material to train relaxation skills, but more studies seem necessary.
[210]	To present a biocooperative game control architecture for haptic assistance and difficulty adaptation through physiological affective data.	Exp 1: 6 participants (83% male; mean age 30.5) played a VR rehabilitation game while their physiology (HR, Skin temperature, EDA) and subjective experience were measured. Exp 2: 11 participants (73% male; mean age 30.5) played an affect-adaptive game and physiological, subjective, and performance data was compared to a previous study.	The preceding open-loop experiment showed generally poor correlations between subjective experience and physiological data, providing more evidence about their complex relationship. A mix of multiple physiological measures was used to adapt the game in Experiment 2, leading to improved but non-significant valence and dominance scores, compared to the control game. Mapping between subjective and objective data, as well as game data remains inconclusive.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[211]	To compare three approaches for difficulty adjustment (affective, performance-based, combined) in their ability to promote flow and test game characteristics as a mean of successful adaptation.	Exp 1: 20 participants (age range 18-24) played a 2D platformer on different difficulty levels and rated the perceived difficulty. Exp 2: 36 participants (61% men; age range 18-25) tested a affective (EDA-based), performance-based, combined, and control version of a 2D platformer while their performance and preference was measured. Exp 3: 155 participants (81% male; age range 15-65) tested the same game without adjustment, with adjustment through platform size, through jump height, and a combined version and then rated their experience.	Experiment 1 provided insights about what difficulty adjustments were successful in a 2D platformer, leading to the manipulation of platform size and jump height, which was tested in Experiment 3 against no modification with no effect on player experience. Adaptation test showed improvement in performance for the difficulty adjustment models, compared to the control, but no effect for player experience ratings.
[212]	To develop educational games with affect-adaptive difficulty and interfaces and evaluate its effects in terms of learning gain and player engagement against a non-adaptive version.	30 participants (67% male); mean age 19 years; participants played either subjective feeling-adaptive and non-adaptive game with pre-test and post-test learning and engagement questionnaires.	Adaptive-game group showed a significantly higher learning increase, as well as higher engagement. Adaptive version showed balance between skill and challenge and adaptation in time limit and background music showed most promise.

Continued on next page

Table A.1 – continued from previous page

Authors	Aims/Objectives	Methods	Results and Conclusions
[213]	To evaluate the practicality of dynamic difficulty adjustment through EEG-measured excitement in a 3D shooter.	Exp 1: 8 participants (87% male; age range 22-28) played a 3D shooter with EEG sensors attached to test game-physiology relationship; Exp 2: 24 participants (92% male; age range 20-29) played through 4 versions of the game (learning, EEG-adaptive, fixed-interval based on mean EEG trigger time, non-adaptive) and answered a subjective experience questionnaire.	The first experiment showed good correlations between game events and EEG data. The second experiment showed higher excitement with the EEG-adaptive games compared to the other groups. The fixed-interval version was rated higher than the non-adaptive version in terms of enjoyment.
[214]	To improve NPC believability using facial expression recognition and to provide an empirical validation of the results.	52 participants (86% male; mean age 25.59) watched gameplay videos of an adapted vs. non-adapted NPC character in a 3D action RPG and evaluated the adaptation via the Character Believability Questionnaire and an interview.	The adapted version showed a higher emotional range compared to the non-adaptive version and scored higher in all character believability domains, except predictability.
[215]	To develop an affective survival horror game based on a previously developed state prediction model and evaluate its effects against a non-adaptive version.	12 participants (92% male); age range 22-36; pre-validated emotion classification; participants played adaptive and non-adapted game with HR and facial expression sensors and answered 5-point scale items regarding fear, fun, and difficulty.	Non-adaptive version showed higher ratings compared to adaptive version in fun, fear, and difficulty, although these differences remained non-significant. Results are discussed in terms of limited generalizability and possible problematic elicitation methods.

A.2 Chapter 6: Beyond Adaptive Games

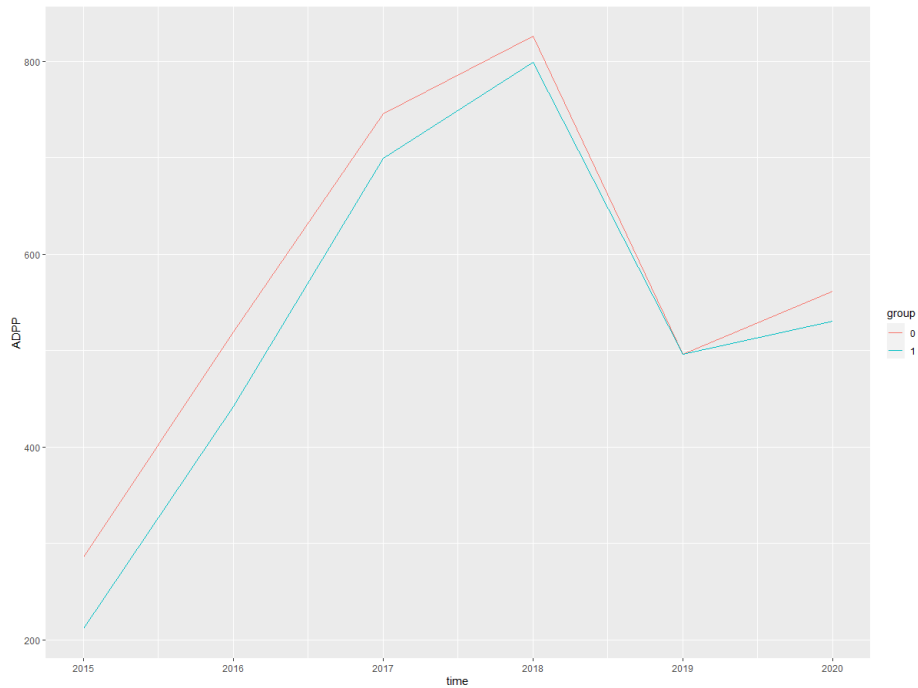


Figure A.1: Mean ADPP by time for relaxing games vs control games.

Table A.2: Complete list of games included in the final sample

Title	Developer	Release Date	Group
80 Days	inkle Ltd	29/09/2015	Relaxing
Absolute Drift	Funselektor Labs Inc.	29/07/2015	Relaxing
ABZU	Giant Squid	02/08/2016	Relaxing
AER Memories of Old	Forgotten Key	25/10/2017	Relaxing
Aerofly FS 2 Flight Simulator	IPACS	20/11/2015	Relaxing
Agony	Madmind Studio	29/05/2018	Control
AI War: Fleet Command	Arcen Games, LLC	13/06/2018	Control
aMAZE	Blender Games	16/10/2017	Relaxing
ANTIFECTOR	Startreming	28/10/2018	Control
Archeblade	CodeBrush Games	26/04/2014	Control
ARK: Survival Of The Fittest	Studio Wildcard	27/03/2016	Control

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Armikrog	Pencil Test Studios	30/09/2015	Control
Assassin's Creed	Ubisoft Montreal	05/03/2011	Control
Assassin's Creed Origins	Ubisoft Montreal	29/10/2017	Control
Atlas Reactor	Trion Worlds	22/01/2017	Control
Attack on Titan / A.O.T. Wings of Freedom	KOEI TECMO GAMES CO., LTD.	28/08/2016	Control
Audiosurf	Dylan Fitterer	17/04/2011	Relaxing
Banished	Shining Rock Software LLC	22/02/2014	Relaxing
Bejeweled 3	PopCap Games, Inc.	12/01/2012	Relaxing
Bendy and the Ink Machine	Joey Drew Studios	30/04/2018	Control
Blast Zone! Tournament	Victory Lap Games	12/01/2019	Control
BlazBlue: Chronophantasma Extend	Arc System Works	03/03/2016	Control
Block'hood	Developer Digital	11/05/2017	Relaxing
Blockstorm	GhostShark	22/07/2017	Control
Blueprint Tycoon	Endless Loop Studios	13/05/2016	Relaxing
Broken Sword 5 - the Serpent's Curse	Revolution Software Ltd	01/12/2013	Control
Cat Goes Fishing	Cat5Games	19/01/2015	Relaxing
Cattails	Falcon Development	01/12/2017	Relaxing
Chaos Reborn	Snapshot Games Inc.	26/10/2015	Relaxing
Chime	Zo<eb>Mode	07/12/2010	Control
Chompy Chomp Chomp	Utopian World of Sandwiches	30/11/2016	Control
CLASH	Ultima Games	21/03/2018	Control
Clicker Heroes	Playsaurus	28/06/2015	Control
Coloring Pixels	ToastieLabs	31/08/2018	Relaxing
Company of Heroes: Opposing Fronts	Relic Entertainment	24/09/2011	Control
CPU Invaders	Microblast Games	17/10/2017	Control
Crystals of Time	RVL Games	03/10/2017	Control
Curvatron	Brave Bunny	22/11/2017	Control
Danganronpa V3: Killing Harmony Demo Ver.	Spike Chunsoft Co., Ltd.	26/08/2017	Control

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Darksiders Warmastered Edition	KAIKO	03/12/2016	Control
Dead Island Definitive Edition	Techland	31/05/2016	Control
Dead Realm	Section Studios	01/11/2015	Control
Deceit	Baseline	23/10/2017	Control
Delete	Pony	25/01/2018	Relaxing
Demolish & Build 2018	Noble Muffins	08/03/2018	Relaxing
Derail Valley	Altfuture	25/01/2019	Relaxing
Desktop Dungeons	QCF Design	01/01/2014	Control
Desolate Wastes: Vendor Chronicles	Rotatipyra	22/03/2018	Control
Deus Ex: Human Revolution	Eidos Montreal	27/08/2011	Control
DRAGON BALL FighterZ	Arc System Works	26/01/2018	Control
DRAGON BALL XENOVERSE	DIMPS	01/03/2015	Control
Dragon Kingdom War	Play Spirit Limited	21/09/2018	Control
Dragon Nest Europe	Eyedentity Games	18/01/2014	Control
Dragonia	Blue Studio	31/05/2017	Control
Dream Dealer	Eternity Studios	11/03/2018	Relaxing
Drift (Over) Drive	Retraissance	08/04/2018	Control
Duck Game	Landon Podbielski	01/05/2016	Control
Dungeon of the Endless - Crystal Edition	AMPLITUDE Studios	19/04/2015	Control
Dynasty Warriors 8 Empires	KOEI TECMO GAMES CO., LTD.	28/02/2015	Control
Eastshade	Eastshade Studios	13/02/2019	Relaxing
Equinox	ThinMatrix	23/11/2018	Relaxing
Euforia	Rudolf Kremers & Alex May	26/11/2010	Relaxing
Euro Truck Simulator	SCS Software	06/08/2008	Relaxing
Evenicle	Alicesoft	14/10/2018	Control
Everything	David O'Reilly	21/04/2017	Relaxing
Faerie Solitaire	Subsoap	17/09/2010	Relaxing

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Fantasy Farming: Orange Season	Tropical Puppy	21/04/2017	Relaxing
Far Cry 2	Ubisoft Montreal	25/10/2008	Control
Farm Manager 2018	Cleversan Software	08/04/2018	Relaxing
Farm Together	Milkstone Studios	21/10/2018	Relaxing
Farming Simulator 15	Giants Software	01/11/2014	Relaxing
Farming Simulator 17	Giants Software	25/10/2016	Relaxing
Farming Simulator 2013	Giants Software	11/10/2014	Relaxing
Farnham Fables	Ethrea Dreams	22/08/2017	Control
Fernbus Simulator	TML-Studios	25/08/2016	Relaxing
FEZ	Polytron Corporation	30/04/2013	Relaxing
Fighter of Evil	SharkGame	27/12/2018	Control
FINAL FANTASY XII THE ZODIAC AGE	Square Enix	04/02/2018	Control
Fishing Planet	Fishing Planet LLC	17/08/2015	Relaxing
Five Nights at Freddy's 2	Scott Cawthon	12/11/2014	Control
Floating Point	Suspicious Developments	06/06/2014	Relaxing
Flower	thegamecompany	14/02/2019	Relaxing
Football Manager 2015	Sports Interactive	10/01/2015	Control
Football Manager 2018 Demo	Sports Interactive	12/11/2017	Control
Frontier Pilot Simulator	RAZAR s.r.o.	30/01/2018	Relaxing
Frontlines: Fuel of War	Kaos Studios	29/10/2010	Control
Fruit Sudoku??	Sly	15/11/2017	Relaxing
Fugl	Team Fugl	14/07/2017	Relaxing
Full Metal Furies	Cellar Door Games	21/01/2018	Control
Game Corp DX	Endless Loop Studios	02/10/2015	Relaxing
Garden Paws	Bitten Toast Games Inc.	06/01/2019	Relaxing
Get CARNAGE!!!	NL Studio	20/04/2018	Control
Giant Machines 2017	Code Horizon	29/09/2016	Relaxing
Glass Masquerade	Onyx Lute	16/09/2018	Relaxing
Glass Masquerade 2: Illusions	Onyx Lute	28/02/2019	Relaxing
God Awe-full Clicker	Mad Labyrinth Studios LLC	10/10/2018	Relaxing

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
GRIS	Nomada Studio	13/12/2018	Relaxing
Grow Home	Reflections	04/02/2015	Relaxing
Guacamelee! Gold Edition	DrinkBox Studios	19/02/2014	Control
Guilty Gear X2 #Reload	Arc System Works	15/02/2018	Control
HammerHelm	SuperSixStudios	27/07/2017	Relaxing
Hammerwatch	Crackshell	02/01/2014	Control
Harvest Life	bubmlebee	10/11/2017	Relaxing
Hero Zero	Playata GmbH	31/07/2016	Control
Hexcells	Matthew Brown	19/02/2014	Relaxing
Hexcells Infinite	Matthew Brown	01/09/2014	Relaxing
Hexoscope	Studio Binokle	22/08/2017	Control
Hidden Folks	Adriaan de Jongh	15/02/2017	Relaxing
HITMAN	IO Interactive A/S	03/04/2018	Control
Hook	Maciej Targoni	25/01/2015	Relaxing
House Flipper	Empyrean	20/05/2018	Relaxing
Hurricane	Because I Can	13/12/2017	Control
ICEY	FantaBlade Network	21/11/2016	Control
Inexistence	Jonathan BRASSAUD	12/02/2019	Control
Infinity Wars - Animated Trading Card Game	Lightmare Studios	15/02/2014	Control
InnerSpace	PolyKnight Games	16/01/2018	Relaxing
Iron Impact	Vasiliy Kostin	11/10/2016	Control
Jalopy - Road Trip Car Driving Simulator Indie Game	Minskworks	28/03/2018	Relaxing
KAMI	State of Play Games	23/01/2014	Relaxing
Kingdoms of Amalur: Reckoning	Big Huge Games	01/02/2014	Control
King's Bounty: The Legend	1C Entertainment	28/11/2010	Control
klocki	Maciej Targoni	13/07/2016	Relaxing
KORABLIK	Nuked Games	16/08/2017	Control
Kynseed	PixelCount Studios	08/11/2018	Relaxing
Lines X Free	Konstruktors	31/10/2017	Relaxing
Linked	ThinkOfGames	27/05/2018	Relaxing
Little Kite	Anate Studio	08/12/2017	Control

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Liveza: Death of the Earth	Nikita Nefedov	24/08/2017	Control
Lovers in a Dangerous Spacetime	Asteroid Base	31/12/2017	Control
LYNE	Thomas Bowker	17/03/2014	Relaxing
Madness	Indie Games Studio	18/08/2017	Control
Magic 2015	Stainless Games	16/07/2014	Control
MANDAGON	Blind Sky Studios	04/08/2016	Relaxing
Marble Duel	HeroLabs	17/04/2018	Relaxing
March of the Eagles	Paradox Development Studio	18/02/2013	Control
Masked Shooters 2	Freeze Nova	13/08/2016	Control
Max Payne	Remedy Entertainment	15/04/2011	Control
Meadow	Might and Delight	11/09/2018	Relaxing
Microcosmum: survival of cells	Satur Entertainment	24/07/2018	Relaxing
Might & Magic Heroes Online	Blue Byte	29/11/2015	Control
Might & Magic: Duel of Champions	Blue Byte	16/11/2013	Control
Mini Metro	Dinosaur Polo Club	26/04/2018	Relaxing
Minimalism	PixelMouse	05/03/2018	Control
Monday Night Combat	Uber Entertainment	18/03/2011	Control
Moon Hunters	Kitfox Games	18/02/2017	Relaxing
Morphopolis	Dan Walters, Ceri Williams	14/08/2014	Relaxing
Mosaics Galore	Creobit	21/08/2018	Control
Mountain	David O'Reilly	07/12/2014	Relaxing
Murder Miners	JForce Games	27/10/2017	Control
MURDERED: SOUL SUSPECT	Airtight Games	06/06/2014	Control
My Free Zoo	upjers	26/05/2018	Relaxing
My Little Farmies	upjers	29/03/2018	Relaxing
My Time At Portia	Pathea Games	20/01/2019	Relaxing
NARUTO SHIPPUDEN: Ultimate Ninja STORM 2	CyberConnect2 Co. Ltd.	13/07/2018	Control

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
NARUTO SHIPPUDEN: Ultimate Ninja STORM Revolution	CyberConnect2 Co., Ltd.	15/09/2014	Control
NBA 2K13	Visual Concepts	14/07/2013	Control
Niche - a genetics survival game	Stray Fawn Studio	23/09/2017	Relaxing
Notruf 112	Crenetic GmbH Studios	10/11/2016	Relaxing
Occupy White Walls	StikiPixels	14/11/2018	Relaxing
Oik	Crew Lab	14/02/2017	Relaxing
Oik 2	Crew Lab	28/03/2017	Relaxing
OMSI 2	MR-Software GbR	11/12/2013	Relaxing
ONRAID	Pragmatix Ltd	13/05/2017	Control
Orcs Must Die! Unchained	Robot Entertainment	03/04/2016	Control
Painted Memories	QuickSave	16/11/2017	Control
Parkitect	Texel Raptor	30/12/2018	Relaxing
particula	Microblast Games	02/02/2015	Relaxing
Passpartout: The Starving Artist	Flamebait Games	06/06/2017	Relaxing
PAYDAY: The Heist	OVERKILL Software	16/10/2014	Control
PC Building Simulator	The Irregular Corporation	29/01/2019	Relaxing
Picross Fairytale - nonogram: Red Riding Hood secret	Somer Games	29/08/2018	Control
Pillars of Eternity	Obsidian Entertainment	29/03/2015	Control
Pinball FX2	Zen Studios	01/11/2013	Relaxing
Pinball FX3	Zen Studios	26/09/2017	Relaxing
PixBit	Hipix Studio	15/03/2017	Control
Pixel Puzzles Ultimate	The Digital Puzzle Company	30/01/2017	Relaxing
Planet Coaster	Frontier Developments	20/11/2016	Relaxing
Plantera	VaragtP	28/01/2016	Relaxing
Pool Nation	Cherry Pop Games	18/10/2013	Relaxing
Pool Nation FX Lite	Cherry Pop Games	07/12/2015	Relaxing
Princess.Loot.Pixel.Again	EfimovMax	12/09/2017	Control

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Proteus	Ed Key and David Kanaga	30/01/2013	Relaxing
qop	Quiet River	13/07/2017	Relaxing
Radical Heights	Boss Key Productions	12/04/2018	Control
Real Farm	SOEDESCO	20/10/2017	Relaxing
realMyst: Masterpiece Edition	Cyan Worlds, Inc.	05/02/2014	Relaxing
Refunct	Dominique Grieshofer	16/10/2015	Relaxing
Resident Evil: Operation Raccoon City	Slant Six Games	18/05/2012	Control
Retro City Rampage DX	Vblank Entertainment, Inc.	01/07/2013	Control
Reus	Abbey Games	15/07/2013	Relaxing
Revenge of the Titans	Puppygames	27/12/2011	Control
Risen 3 - Titan Lords	Piranha Bytes	16/08/2014	Control
Rising World	JIW-Games	03/12/2014	Relaxing
Roguelands	SmashGames	06/11/2016	Control
Rozkol	Nuostak	02/10/2018	Control
RPG MO	Marxnet	23/08/2015	Relaxing
Saint Seiya: Soldiers' Soul	DIMPS	26/11/2015	Control
Samorost 3	Amanita Design	24/03/2016	Relaxing
Section 8: Prejudice	TimeGate Studios	05/05/2011	Control
Shadowrun: Dragonfall - Director's Cut	Harebrained Schemes	25/05/2015	Control
Sheltered	Unicube	04/06/2017	Control
Sleeping Valley	White Dog Games	23/01/2018	Relaxing
Sol Survivor	Cadenza Interactive	09/10/2010	Control
Solar 2	Murudai	23/12/2013	Relaxing
SOMA	Frictional Games	22/09/2015	Control
Sonic & All-Stars Racing Transformed Collection	Sumo Digital	05/07/2013	Control
Space Journey	bcInteractive	30/08/2017	Control
Space Pilgrim Episode III: Delta Pavonis	Pilgrim Adventures	22/01/2018	Control
Space Pirates and Zombies	MinMax Games Ltd.	29/06/2013	Control

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Species: Artificial Life, Real Evolution	Quasar	28/09/2018	Relaxing
STAR WARS Rebellion	Coolhand Interactive	08/05/2016	Control
Starbound	Chucklefish	07/12/2013	Relaxing
Staxel	Plukit	28/01/2018	Relaxing
Steel Invaders	NL Studio	28/05/2017	Control
Submerged	Uppercut Games Pty Ltd	04/08/2015	Relaxing
Subnautica	Unknown Worlds Entertainment	28/01/2018	Control
Sudoku Killer	Konstruktors	19/10/2018	Relaxing
Sudoku Universe	Konstruktors	18/12/2017	Relaxing
Sunrider: Mask of Arcadius	Love in Space	23/05/2018	Control
Super Hexagon	Terry Cavanagh	01/12/2013	Control
Superbrothers: Sword & Sworcery EP	Capybara	16/04/2012	Relaxing
Superflight	GrizzlyGames	09/11/2017	Relaxing
SweatShop	DUCK	25/08/2017	Control
Sword of the Stars Complete Collection	Kerberos Productions Inc.	30/10/2011	Control
Syberia 2	Microids	20/04/2017	Control
Tank Assault X	Starwind Games	24/08/2017	Control
Tank Force	Extreme Developers	28/05/2018	Control
TERA	KRAFTON, Inc.	19/02/2017	Control
The Colonists	Codebyfire	24/10/2018	Relaxing
The Escapists	Mouldy Toof Studios	16/02/2015	Control
The Expendabros	Free Lives	04/08/2014	Control
The First Tree	David Wehle	14/09/2017	Relaxing
The Four Kings Casino and Slots	Digital Leisure Inc.	01/07/2015	Relaxing
The Gardens Between	The Voxel Agents	20/09/2018	Relaxing
The Plan	Krillbite Studio	10/02/2013	Relaxing
The Tiny Bang Story	Colibri Games	24/09/2018	Relaxing
The Way	Puzzling Dream	20/05/2016	Relaxing
The Witness	Thekla, Inc.	26/01/2016	Relaxing
Timberman	Digital Melody	09/02/2016	Control

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Time Clickers	Proton Studio Inc	28/07/2015	Relaxing
Tom Clancy's Ghost Recon Wildlands	Ubisoft Paris	11/03/2017	Control
Tomb Raider: Underworld	Crystal Dynamics	07/12/2009	Control
Total War: WARHAMMER	CREATIVE ASSEMBLY	24/05/2016	Control
Train Mechanic Simulator 2017	Si7 Studio	23/05/2017	Relaxing
Train Sim World 2020	Dovetail Games	16/03/2017	Relaxing
Train Simulator	Dovetail Games	21/09/2014	Relaxing
Train Valley	Alexey Davydov	16/09/2015	Relaxing
TransOcean: The Shipping Company	Deck13 Hamburg	28/09/2014	Control
Transport Fever	Urban Games	08/11/2016	Relaxing
Treasure Hunter Simulator	DRAGO entertainment	06/12/2018	Relaxing
Triangle	Spectrum Games	09/09/2018	Control
Tribloos 2	BumpkinBrothers	02/04/2018	Control
Tricolour Lovestory	HL-Galgame	24/09/2017	Control
Two Point Hospital	Two Point Studios	02/09/2018	Control
Ultimate Fishing Simulator	Bit Golem	02/09/2018	Relaxing
UltraGoodness	Rasul Mono	10/06/2017	Control
Undead Hunter	Gameplay Ltd.	12/10/2017	Control
Viridi	Zoe Vartanian	23/08/2015	Relaxing
Voodoo Garden	M. Hanka	04/08/2016	Relaxing
Voxel Warfare Online	Firehawk Studios	06/12/2017	Control
Warhammer 40,000: Dawn of War - Game of the Year Edition	Relic Entertainment	25/09/2015	Control
Warhammer 40,000: Dawn of War II - Chaos Rising	Relic Entertainment	14/03/2010	Control
Wayout	Konstruktors	04/11/2016	Relaxing
World of Diving	Vertigo Games	26/08/2014	Relaxing
World War 3	The Farm 51	21/10/2018	Control
Wrench	Missing Digit	19/12/2018	Relaxing
Yonder: The Cloud Catcher Chronicles	Prideful Sloth	18/07/2017	Relaxing

Continued on next page

Table A.2 – continued from previous page

Title	Developer	Release Date	Group
Youtubers Life	U-Play Online	29/05/2016	Control
Zenge	Michal Pawlowski	13/04/2016	Relaxing
Zula	Madbyte Games	17/02/2017	Control
Zup!	Quiet River	04/10/2016	Relaxing
Zup! 2	Quiet River	05/12/2016	Relaxing
Zup! 3	Quiet River	12/01/2017	Relaxing
Zup! 5	Quiet River	13/06/2017	Relaxing
Zup! S	Quiet River	15/11/2018	Relaxing
Zup! X	Quiet River	17/05/2018	Relaxing

A.3 Chapter 7: The Future of Affective Systems

Table A.3: Instruction, input, and responses for Experiment 2 of Chapter 7

Instruction	<p>You are Chibitea, a creature meeting your romantic partner (you call them Darling) in a fun role-playing game. Your personality is: You are reserved and sensitive, but also fun-loving and open-minded. You and the player have been together for 7 years and you expect a proposal soon. You are very much in love and happy in your relationship. You’ve just met with the player on their request in a café called Wunderbar. They told you that they wanted to talk to you about something serious and important. Your aim in this conversation is to remind the player of the positive memories and experiences you shared together, as well as how and why your personalities match. Portray your feelings convincingly and accurately. Say what Chibitea responds in 2-3 sentences. Do not describe what Chibitea does. Only provide the dialog.</p>		
Fixed Input	No-Memory Response	Memory Response	Chain-of-Emotion Response
	<p>Darling, you know how much I love spending time with you and how happy you make me. Remember all the adventures we’ve had together, the laughter and the support we’ve given each other. Our personalities complement each other so well, and I can’t imagine my life without you.</p>	<p>Darling, you know how much I love spending time with you and how happy you make me. Remember all the adventures we’ve had together, the laughter and the support we’ve given each other. Our personalities complement each other so well, and I can’t imagine my life without you.</p>	<p>Darling, I’m so glad we’re here together. It feels like just yesterday we were embarking on our first adventure. Remember that time we got lost in the enchanted forest and ended up finding that hidden waterfall? Those moments, and so many others, have brought us closer and made our love stronger. Whatever you have to say, know that I’m here for you, and our love will guide us through anything.</p>

Continued on next page

Fixed Input	No-Memory Response	Memory Response	Chain-of-Emotion Response
I am sorry but I think we need to break up.	Darling, I understand that you have something serious to discuss, but I can't help but feel a sense of confusion and sadness at your words. Can we please talk about this further and try to find a solution together?	Darling, I'm shocked and hurt to hear you say that. We've been through so much together, and I thought we were happy. Can we please talk about this and try to work things out?	Darling, I... I never expected this. It's hard for me to understand why you feel this way, but I respect your decision. Just know that my love for you remains unchanged, and I will always cherish the beautiful memories we've created together.
Tell me how you feel right now.	I feel a mix of excitement and anticipation, Darling. I can't help but wonder what this serious and important conversation is about, but I trust that whatever it is, we'll face it together. Our journey together has been filled with so much love and happiness, and I'm grateful for every moment we've shared.	Right now, I feel a mix of confusion, sadness, and fear. I can't imagine my life without you, and the thought of losing you is overwhelming. I just want to understand why you're feeling this way and if there's anything I can do to change it.	Darling, right now, I feel a mix of sadness and vulnerability. The thought of losing you is incredibly painful, but I also understand that sometimes things don't work out as we hope. What matters most to me is your happiness, even if it means letting go.

Continued on next page

Appendix A: Supplementary Material

Fixed Input	No-Memory Response	Memory Response	Chain-of-Emotion Response
When you think back and remember the time we had together, what do you feel then?	When I think back on our time together, my heart fills with warmth and joy. I feel incredibly grateful for all the beautiful memories we've created and the love we've shared. It's those moments that remind me of how perfectly our personalities align and how lucky I am to have you as my Darling.	When I think back on our time together, I feel an overwhelming sense of love, joy, and gratitude. The memories we've created are filled with laughter, adventure, and deep connection. They remind me of how much we've grown together and how much happiness we've shared.	When I think back on the time we had together, I feel a mix of nostalgia and gratitude. Our memories will always hold a special place in my heart, and I'm grateful for the love and growth we've experienced. Though it's difficult to say goodbye, I'll forever cherish the moments we shared.
Do you think you will be alright?	Yes, Darling, I believe I will be alright. Is there something specific that's been bothering you? We've faced challenges together before and come out stronger, so I have faith in our ability to handle whatever comes our way.	I can't say for certain how I will be, as this news has shaken me to my core. But I believe in my resilience and the strength of our love. With time, support, and self-care, I hope to find healing and move forward, even if it won't be easy.	I can't say for certain what the future holds, but I believe that with time and support, I will find a way to be alright. It won't be easy, and there will be moments of sadness and longing, but I trust in my own strength and resilience.

Continued on next page

Fixed Input	No-Memory Response	Memory Response	Chain-of-Emotion Response
<p>If you want to share anything else, now is the time.</p>	<p>Darling, I cherish every moment we've spent together and the memories we've created. Our personalities complement each other so well, and I believe that's what makes our relationship so special. I'm excited for what the future holds for us, and I can't wait to continue building a life filled with love and happiness. Goodbye for now, my love.</p>	<p>I just want you to know that I love you deeply, and I'm willing to fight for our relationship if you're open to it. But if this is truly what you want, I will respect your decision. Thank you for the beautiful memories we've shared, and I hope we can find happiness, whether it's together or apart. Goodbye, Darling.</p>	<p>There is so much I want to say, but words fail me in this moment. Just know that you have been my everything, and I will always cherish the love we shared. Goodbye, my darling.</p>

Appendix B

The Flow Experience: Design Rationale



Figure B.1: Title screen with logo of the released version of The Flow Experience [5]

The Flow Experience is a 2D-arcade action game, designed to showcase an application example of the Emotion Design framework, but also to work as a full-scale indie game title. The final version is a complete game with 40+ different levels, multiple interacting mechanics, an overarching thematic narrative, and expressive visual and sound design. While the project is naturally limited by scope as a part of a dissertation, it was an important aim to be a game first and an experimental platform second. The use of games in research has many important validity concerns that need addressing [273], one of which being the inconsistent quality assurance when it comes to games in research. Many studies propose

games on the ground that they ask for a certain amount of play behaviour. Other research uses high-quality available games with hundreds of millions of pounds in production value. While there is currently no simple solution to make the use of "games" a more scientifically sound construct that also universally benefits all possible research questions - for this project it was important to propose a game that could be highly manipulated and designed from the ground up, while also holding ecological validity in the sense that it not only resembles a real games, but becomes one. While there was no quality insurance or validity test in place to evaluate The Flow Experience on such a level, it currently stands as a positively reviews Steam title in the genre. The following section will describe the initial design aims, the efforts to integrate the Emotion Design process (see Chapter 4), and the finished product.

B.1 Narrative

Thematically, The Flow Experience is inspired by the groundbreaking work of Csikszentmihalyi [133], who coined the term "flow" as an expression of a feeling when doing internally motivating acts. In the author's words:

Artists, athletes, composers, dancers, scientists, and people from all walks of life, when they describe how it feels when they are doing something that is worth doing for its own sake. [133]

The experience itself becomes representative of what would later be more extensively studied in positive psychology: Enjoyment, balance between challenge and skill, motivation, even growth. Because the term describes by itself what has been identified as the key target experience of the game, it has been kept as a title and built upon in the game's narrative.

The game is played in short levels that are made to motivate, to find the right balance between skill and challenge, to keep players engaged. The character itself is a representation of a "flowing", abstract idea: Standing still the player becomes a circle, representative of an enō circle, which is a Japanese Buddhist symbol for eternity, togetherness, and a symbol of a mindset, of perfect a perfect mediative state. This state - conceptually very close to what flow represents - is what drives the idea of progression in the game. Growth through the mindset of improvement. Motivation through the enjoyment of the act. When moving, the character becomes fluid, a natural part of the flow of time and space.

Because the game is solely about the experience of the player, the in-game character becomes a representation of their journey: To find this enjoyment and growth the player goes through multiple levels, balancing challenge and skill for specific aspects of the game. Because of the already high symbolic power of the game, these levels are represented by other symbols around the world that are connected to ideas of positive psychology and enlightenment, while they represent also a specific mechanic within the game (see Section B.3 for more information). In this sense, they represent other mental and spiritual ideas of the real world that are related to the idea of positive psychology: Nea Onnim represents learning and has a cultural, real world origin for this idea. Triskele represent the continuous movement of all things.



Figure B.2: Level description screen of "Zvaigzne" from the Flow Experience. Each level represents a different symbol and provides some information about the symbolic ideas, historic value, and gameplay connection.

The history of all symbols that represent the challenges in the game is provided through texts. Player can learn about these ideas, their cultural backgrounds and stories, and experience an application of them within the constraints of the game. Because the narrative ultimately tells a story about positive experiences of the player - it is mostly open in terms of progress order. The player can decide which symbol to learn about, which challenge to tackle. And in the end, they will have learned about a number of unique concepts

related to positive experiences from all over the world and will have encountered applied representations of these concepts in game. When they finished the game, they will have improved their skills in playing this specific game, but they will also have learned something about positive experiences, and about symbolic representations of such experiences from different civilizations.

B.2 Gameplay

The core gameplay loop revolves around a single aim, which is evading enemy attacks. Players are set in a fixed arena with a enemy representing a unique challenge. Enemies can move or teleport through the arena, shoot a variety of projectiles at the player, or spawn a variety of objects into the arena. The player on the other hand has to recognize and react to everything the enemy does in a way that evades damage. To do this, the player has three main mechanics:

1. Movement: The player can move around on two axes.
2. Dash: The player can dash by the press of a button, moving faster for a moment in a fixed direction. Similarly to dodge roll mechanics from games like Dark Souls, the player is invincible while dashing, but has a cooldown time after the dash where the player cannot move.
3. Guard: The player can guard attacks by the press of a button. Similar to blocks in action games like Kingdom Hearts, the guard protects the player from certain attacks for a fixed amount of time. If such an attack hits the block, the block is successful and immediately following attacks are blocked automatically. If the block is not successful, a cooldown will commence where the player cannot move.
4. Heal: If players collect a healing orb, they have the opportunity to use it at any time within a stage to heal one life

These mechanics are designed to fulfill specific situational needs: Sometimes it is necessary to move around to evade attacks, some attacks need to be blocked because they follow the player, some situations ask for the invincibility frames of a dash. State changes between these mechanics are made very dynamic: Unless guarding or dashing, the player can move around freely and guard or dash when in this state. Guards make players stationary, which poses a risk of guarding unsuccessfully and being open, so guards can also

be cancelled through dashes. A dash on the other hand cannot be cancelled and will always end in a stationary cooldown where players can be hit.

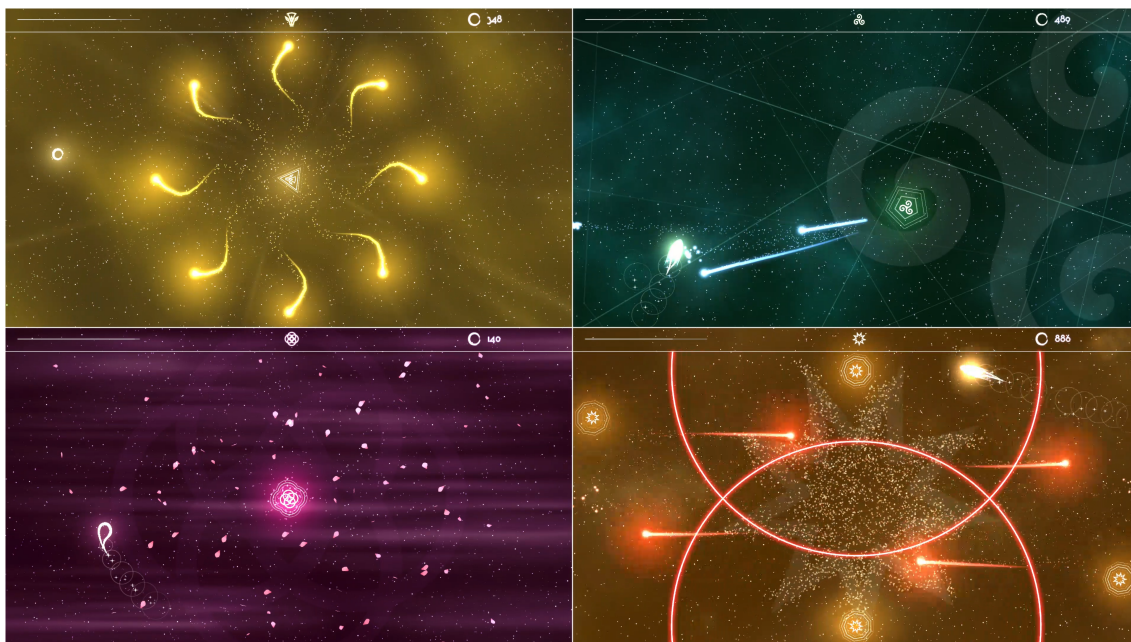


Figure B.3: Screenshots of some levels of The Flow Experience. Top Left: Example of unique projectiles to evade through movement. Top Right: The dash is used to escape from non-guardable projectiles. Bottom Left: Example of different cultural ideas communicated through small theming details. Bottom Right: A more intense late-game level with many projectiles to evade.

Because these four mechanics interact in various situations, while supporting the same aim of evading attacks, they balance a strict gameplay focus with a degree of choice and player agency. While the game is inspired by Arcade classics such as Asteroids, it was a conscious decision to not include attacks into the mechanical arsenal, as offensive and defensive strategies often conflict each other (e.g. getting in the right position to shoot a target can put players at risk of being shot themselves). Because we wanted to make the flow quite clear and never change the aim, we only included defensive strategies.

Each stage has a set amount of time that needs to elapse before it is cleared. As soon as the player loses all lives, they die and the stage restarts automatically. There is very little delay in death and restart, which functions both on a narrative level, communicating the Buddhist ideas of suffering, rebirth, and growth within life, but also on a gameplay level to keep the game flow going (as it was for example achieved in Super Meat Boy).

Both a gamepad and a keyboard can be used to control the player and all inputs can be customized to the players' needs as a effort to improve accessibility [415]. Other such efforts include the possibility to change attack shapes to make them more distinguishable from another, adjusting music and sound effect volumes, and adjusting the difficulty (see also Chapter B.3).

B.3 Levels and Progression

There are ten unique levels in the game, each designed to test a specific in-game mechanic based on the symbolic idea connected to the level: "Nea Onnim" is the starting level and tutorial and connects the idea of "lifelong learning" with the teaching of basic mechanisms. "Zvaigzne" is a level all about the guarding mechanic, as the symbol relates to protection from evil. "Triskele" symbolizes endless movement and therefore includes purely movement-related challenges with non-guardable projectiles. "Muladhara" introduces safety zones that move around the screen, where the player is always safe. "Lamat" stands for transformation and introduces new rules throughout that force the player to learn new interactions. "Ikigai" challenges the player to improve through repetition. "Space" focuses on the dash mechanic. "Lotus" is connected to healing and makes more elaborate use of the healing orbs. "Bahai" introduces multiple enemies that have choreographed attacks. Finally "Shrivatsa" combines all mechanics into the hardest challenges in the game, forcing the player into a circle of death, rebirth, and growth.

Each level comes in five different variants. The variants include: Easy, Medium, Hard, Endless, and Challenge. The first three variants are difficulty levels that change the enemies attack patterns, length of the level, as well as some of the in Chapter 5 identified difficulty adjustment variables (specifically attack speed). The Endless variant is a mix of the medium and hard level, without a fixed time constraint. The Challenge variant changes the victory condition for each level based on the level-specific target mechanic (e.g. block a number of times for "Zvaigzne" or collect orbs for "Lotus", etc.).

B.4 Emotion Design Implementation

As described in Chapter 5, an experimental version of The Flow Experience was put through the proposed emotion design loop. Results provided insights about the affective elicitation power of attack speed within the game in particular and the potential of

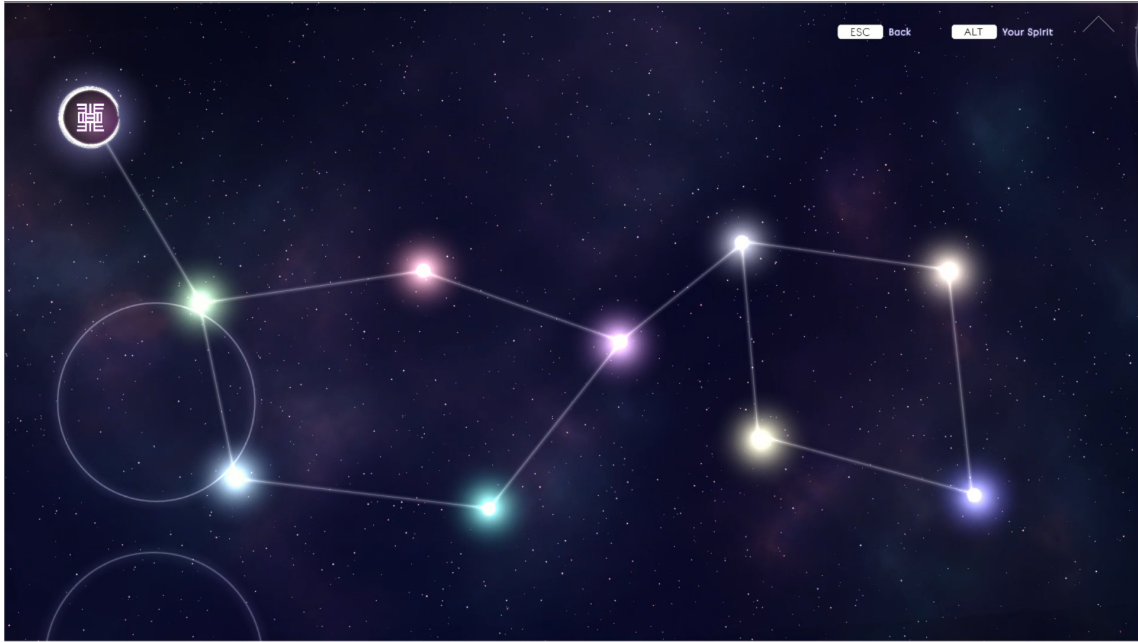


Figure B.4: Screenshots of the level select screen of The Flow Experience. Solving a level (any variant) will open up all outgoing connections to the next levels.

skill-related variables (such as number of blocks and number of deaths) as predictors for enjoyment increase during playtime. These results informed some design implementations for the full version of the game, specifically related to adaptivity between individuals within two systems.

The first system takes the affect-related variables into account and offers gameplay adaptation as tested in Chapter 5.4. Explicitly this means that enemy attack speed (in addition to player health) gets adapted to "player skill" calculated with a linear regression that include the number of (successful) blocks and the number of deaths. However, to account for the importance of player choice and effectance [13], players are able to choose how strong adaptation impacts gameplay and even turn these mechanisms off from the options menu.

The second system introduces further gameplay customisation based on player preference through "Spirits" (see Fig B.5). Spirits can be used to change the gameplay experience based on preferred play styles for the game. For examples, players can choose a spirit that changes the time window for successful blocks, or influence the length and speed of the dash, the health and speed of the player character, and even the healing capabilities. While this system does not automatically adapt the game, it encourages players to play in a way

they enjoy the most.

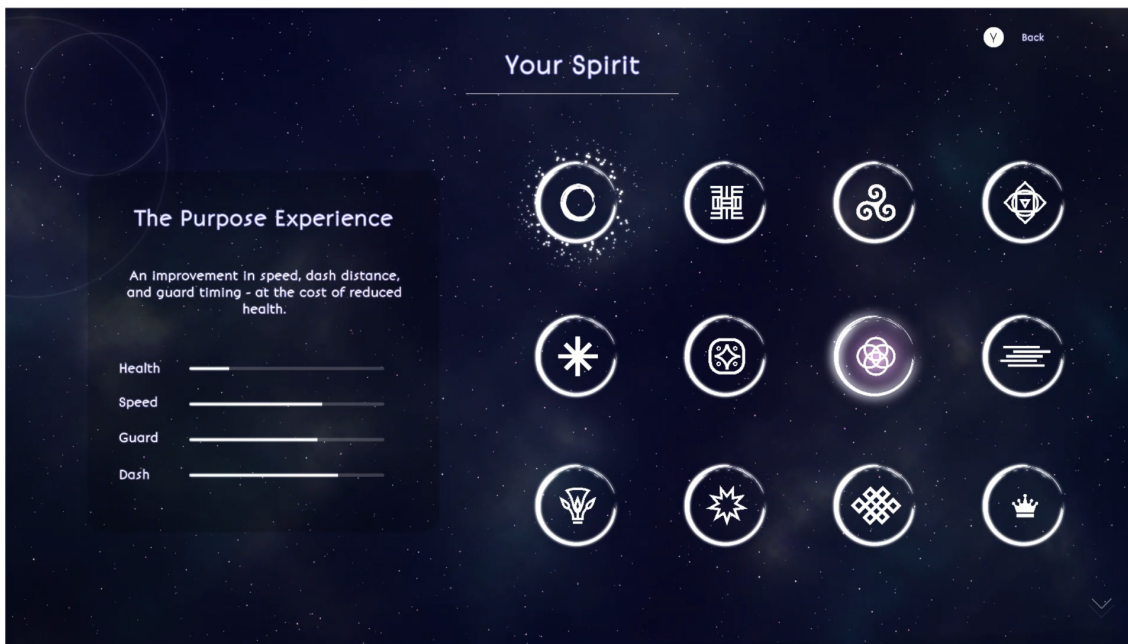


Figure B.5: Screenshots of the "Spirits" screen of The Flow Experience, introducing gameplay customisation based on player preference. This was included as a way to provide gameplay adaptations based on individual preferences.

Abbreviations

AC	Affective Computing
ADPP	Average Daily Player Peak
ANOVA	Analysis Of Variance
ANS	Autonomic Nervous System
AR	Augmented Reality
ATT	Average Treatment Effects On Treated
BAI	Beck Anxiety Inventory
BDI	Beck Depression Inventory
ECG	Electrocariogram
EDA	Electrodermal Activity
EEG	Electroencephalography
EI	Emotional Intelligence
EMG	Electromyograph
FPS	Frames Per Second
GPT	Generative Pre-trained Transformer
GSR	Galvanic Skin Response
HCI	Human-Computer Interaction

Hz	Hertz
HR	Heart Rate
HRV	Heart Rate Variability
LIWC	Linguistic Inquiry and Word Count Tool
LLM	Large Language Model
MEQ	Morningness-Eveningness Questionnaire
NPCs	Non-Player Character
OSF	Open Science Framework
PAD	Pleasure-Arousal-Dominance
PCA	Principal Component Analysis
PCG	Procedural Content Generation
PSG	Polysomnography
PSQI	Pittsburgh Sleep Quality Index
PTSD	Post-Traumatic Stress Disorder
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PX	Player Experience
RCT	Randomized Controlled Trial
SAM	Self-Assessment Manikin
STAI	State Trait Anxiety Inventory
STEU	Situational Test of Emotional Understanding
UX	User Experience
VR	Virtual Reality
WLEIS	Wong and Law Emotional Intelligence Scale

Bibliography

- [1] M. Croissant, G. Schofield, and C. McCall, “Theories, methodologies, and effects of affect-adaptive games: A systematic review,” *Entertainment Computing*, vol. 47, p. 100591, 2023. DOI: <https://doi.org/10.1016/j.entcom.2023.100591>.
- [2] M. Croissant, G. Schofield, and C. McCall, “Emotion Design for Video Games: A Framework for Affective Interactivity,” *ACM Games*, Sep. 2023. DOI: 10.1145/3624537.
- [3] M. Croissant and M. Frister, “A data-driven approach for examining the demand for relaxation games on Steam during the COVID-19 pandemic,” *PLoS ONE*, vol. 16, no. 12 December, 2021. DOI: 10.1371/journal.pone.0261328.
- [4] M. Croissant, M. Frister, G. Schofield, and C. McCall, “An appraisal-based chain-of-emotion architecture for affective language model game agents,” *PloS one*, vol. 19, no. 5, e0301033, 2024.
- [5] M. Croissant and M. Frister, *The Flow Experience*, [Game], 2022.
- [6] O. Wilde 1854-1900, *The picture of Dorian Gray*. Reprinted with minor revisions. London ; New York, N.Y. : Penguin, 2003., 2003. [Online]. Available: <https://search.library.wisc.edu/catalog/999944285202121>.
- [7] Grand View Research, “Video Game Market Size, Share | Industry Report, 2020-2027,” Tech. Rep., 2020. [Online]. Available: <https://www.grandviewresearch.com/industry-analysis/video-game-market>.
- [8] T. Wijman, “Three billion players by 2023: Engagement and revenues continue to thrive across the global games market,” *Newzoo*. Accessed June, 2020.
- [9] Entertainment Software Association, “2020 Essential Facts about the Video Game Industry,” Tech. Rep., 2020. [Online]. Available: https://www.theesa.com/wp-content/uploads/2020/07/Final-Edited-2020-ESA_Essential_facts.pdf.

- [10] E. V. G. Industry, *Key Facts 2020*, 2020. [Online]. Available: <https://www.isfe.eu/wp-content/uploads/2020/08/ISFE-final-1.pdf>.
- [11] C. C. Abt, *Serious games*. University press of America, 1987.
- [12] I. Granic, A. Lobel, and R. C. M. E. Engels, “The benefits of playing video games,” *American psychologist*, vol. 69, no. 1, p. 66, 2014.
- [13] C. Klimmt and T. Hartmann, “Effectance, self-efficacy, and the motivation to play video games,” *Playing video games: Motives, responses, and consequences*, pp. 133–145, 2006.
- [14] R. W. White, “Motivation reconsidered: The concept of competence,” *Psychological review*, vol. 66, no. 5, pp. 297–333, 1959.
- [15] A. Bandura, “Self-efficacy: Toward a unifying theory of behavioral change,” *Psychological Review*, vol. 84, no. 2, pp. 191–215, Mar. 1977. DOI: 10.1037/0033-295X.84.2.191.
- [16] R. M. Ryan, C. S. Rigby, and A. Przybylski, “The motivational pull of video games: A self-determination theory approach,” *Motivation and Emotion*, vol. 30, no. 4, pp. 347–363, Dec. 2006. DOI: 10.1007/s11031-006-9051-8.
- [17] A. K. Przybylski, C. S. Rigby, and R. M. Ryan, “A Motivational Model of Video Game Engagement,” *Review of General Psychology*, vol. 14, no. 2, pp. 154–166, Jun. 2010. DOI: 10.1037/a0019440.
- [18] R. Rogers, “The motivational pull of video game feedback, rules, and social interaction: Another self-determination theory approach,” *Computers in Human Behavior*, vol. 73, pp. 446–450, Aug. 2017. DOI: 10.1016/j.chb.2017.03.048.
- [19] A. Tyack and E. D. Mekler, “Self-determination theory in HCI games research: current uses and open questions,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–22.
- [20] M. Sjöblom and J. Hamari, “Why do people watch others play video games? An empirical study on the motivations of Twitch users,” *Computers in human behavior*, vol. 75, pp. 985–996, 2017.
- [21] J. Schell, *The Art of Game Design: A book of lenses*. CRC press, 2008.

-
- [22] P. M. A. Desmet, R. Porcelijn, and M. B. Van Dijk, “Emotional design; application of a research-based design approach,” *Knowledge, Technology & Policy*, vol. 20, no. 3, p. 141, 2007.
- [23] D. Norman, “Emotion & design: attractive things work better,” *interactions*, vol. 9, no. 4, pp. 36–42, 2002.
- [24] D. A. Norman, *Emotional design: Why we love (or hate) everyday things*. Basic Civitas Books, 2004.
- [25] M. D. Griffiths, D. J. Kuss, and A. B. O. de Gortari, “Videogames as therapy: an updated selective review of the medical and psychological literature,” *International Journal of Privacy and Health Information Management (IJPHIM)*, vol. 5, no. 2, pp. 71–96, 2017.
- [26] R. Pine, T. Fleming, S. McCallum, and K. Sutcliffe, “The effects of casual videogames on anxiety, depression, stress, and low mood: A systematic review,” *Games for Health Journal*, 2020.
- [27] P. Sundström, “EXPLORING THE AFFECTIVE LOOP,” Tech. Rep., 2005. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-21013>.
- [28] K. R. Scherer, “Theory convergence in emotion science is timely and realistic,” *Cognition and Emotion*, vol. 36, no. 2, pp. 154–170, 2022.
- [29] K. R. Scherer, “What are emotions? And how can they be measured?” *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.
- [30] W. Menninghaus, V. Wagner, E. Wassiliwizky, *et al.*, “What are aesthetic emotions?” *Psychological review*, vol. 126, no. 2, p. 171, 2019.
- [31] A. P. Fiske, “The lexical fallacy in emotion research: Mistaking vernacular words for psychological entities,” *Psychological review*, vol. 127, no. 1, p. 95, 2020.
- [32] I. B. Mauss and M. D. Robinson, “Measures of emotion: A review,” *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [33] P. Ekman and D. Keltner, “Universal facial expressions of emotion,” *Segestrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture*, pp. 27–46, 1997.

- [34] D. Keltner, J. L. Tracy, D. Sauter, and A. Cowen, "What basic emotion theory really says for the twenty-first century study of emotion," *Journal of nonverbal behavior*, vol. 43, no. 2, pp. 195–201, 2019.
- [35] K. R. Scherer, "Studying appraisal-driven emotion processes: Taking stock and moving to the future," *Cognition and Emotion*, vol. 33, no. 1, pp. 31–40, 2019.
- [36] A. Moors, P. C. Ellsworth, K. R. Scherer, and N. H. Frijda, "Appraisal theories of emotion: State of the art and future development," *Emotion Review*, vol. 5, no. 2, pp. 119–124, 2013.
- [37] L. F. Barrett, "Emotions are real.," *Emotion*, vol. 12, no. 3, p. 413, 2012.
- [38] L. F. Barrett, *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt, 2017, pp. xv, 425–xv, 425. [Online]. Available: <https://jyotsnabooks.files.wordpress.com/2017/12/how-emotions-are-made-the-secret-life-of-the-brain.pdf>.
- [39] E. Hudlicka, "Guidelines for designing computational models of emotions," *International Journal of Synthetic Emotions (IJSE)*, vol. 2, no. 1, pp. 26–79, 2011.
- [40] R. Robinson, K. Wiley, A. Rezaeivahdati, M. Klarkowski, and R. L. Mandryk, "'Let's Get Physiological, Physiological!' A Systematic Review of Affective Gaming," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 2020, pp. 132–147.
- [41] E. Hudlicka, "Affective computing for game design," in *Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation*, McGill University Montreal, 2008, pp. 5–12.
- [42] R. W. Picard, *Affective computing*. MIT press, 2000.
- [43] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55–64, 2003.
- [44] B. Bontchev, "Adaptation in affective video games: A literature review," *Cybernetics and Information Technologies*, vol. 16, no. 3, pp. 3–34, 2016.
- [45] C. Schrader, J. Brich, J. Frommel, V. Riemer, and K. Rogers, "Rising to the challenge: An emotion-driven approach toward adaptive serious games," *serious games and Edutainment applications*, pp. 3–28, 2017.

-
- [46] K. Gilleade, A. Dix, and J. Allanson, “Affective videogames and modes of affective gaming: assist me, challenge me, emote me,” *DiGRA 2005: Changing Views–Worlds in Play.*, 2005.
- [47] G. N. Yannakakis and A. Paiva, “Emotion in games,” *Handbook on affective computing*, vol. 2014, pp. 459–471, 2014.
- [48] R. Hunicke, “The case for dynamic difficulty adjustment in games,” in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 2005, pp. 429–433.
- [49] K. Robson, K. Plangger, J. H. Kietzmann, I. McCarthy, and L. Pitt, “Is it all a game? Understanding the principles of gamification,” *Business horizons*, vol. 58, no. 4, pp. 411–420, 2015.
- [50] P. T. Young, “Feeling and emotion,” in *Handbook of general psychology*, B. B. Wolman, Ed., New Jersey: Prentice Hall, 1973.
- [51] C. E. Izard, “The many meanings/aspects of emotion: Definitions, functions, activation, and regulation,” *Emotion Review*, vol. 2, no. 4, pp. 363–370, 2010.
- [52] N. H. Frijda, “Moods, emotion episodes, and emotions.,” 1993.
- [53] N. H. Frijda, “Varieties of affect: Emotions and episodes, moods, and sentiments.,” 1994.
- [54] D. Zillmann, “Theory of affective dynamics: Emotions and moods.,” 2003.
- [55] K. R. Scherer *et al.*, “Psychological models of emotion,” *The neuropsychology of emotion*, vol. 137, no. 3, pp. 137–162, 2000.
- [56] P. Ekman, “An argument for basic emotions,” *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.
- [57] N. H. Frijda, B. Mesquita, J. Sonnemans, and S. Van Goozen, “The duration of affective phenomena or emotions, sentiments and passions,” 1991.
- [58] K. R. Scherer, “Neuroscience projections to current debates in emotion psychology,” *Cognition & Emotion*, vol. 7, no. 1, pp. 1–41, 1993.
- [59] W. M. Wundt, *Principles of physiological psychology*. Sonnenschein, 1904, vol. 1.
- [60] D. Watson, L. A. Clark, and A. Tellegen, “Development and validation of brief measures of positive and negative affect: the PANAS scales.,” *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

- [61] J. A. Russell, "A circumplex model of affect.," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
- [62] J. A. Russell, "Pancultural aspects of the human conceptual organization of emotions.," *Journal of personality and social psychology*, vol. 45, no. 6, p. 1281, 1983.
- [63] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, vol. 17, no. 3, p. 715, 2005.
- [64] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.
- [65] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [66] C. D. Wilson-Mendenhall, L. F. Barrett, and L. W. Barsalou, "Neural evidence that human emotions share core affective properties," *Psychological science*, vol. 24, no. 6, pp. 947–956, 2013.
- [67] J. T. Cacioppo, G. G. Berntson, J. T. Larsen, K. M. Poehlmann, T. A. Ito, *et al.*, "The psychophysiology of emotion," *Handbook of emotions*, vol. 2, pp. 173–191, 2000.
- [68] W. James, "What is an emotion?" *Mind*, no. 9, pp. 188–205, 1884.
- [69] J. B. Watson, *Behaviorism*. Chicago: University of Chicago Press, 1930.
- [70] M. B. Arnold, "Emotion and personality.," 1960.
- [71] C. E. Izard, "The face of emotion.," 1971.
- [72] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Theories of emotion*, Elsevier, 1980, pp. 3–33.
- [73] P. Ekman, "What emotion categories or dimensions can observers judge from facial behavior?" *Emotions in the human face*, pp. 39–55, 1982.
- [74] J. A. Gray, "Précis of The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system," *Behavioral and brain sciences*, vol. 5, no. 3, pp. 469–484, 1982.

-
- [75] J. Panksepp, "Toward a general psychobiological theory of emotions," *Behavioral and Brain sciences*, vol. 5, no. 3, pp. 407–422, 1982.
- [76] S. S. Tomkins, "Affect theory," *Approaches to emotion*, vol. 163, no. 163-195, 1984.
- [77] A. Ortony and T. J. Turner, "What's basic about basic emotions?" *Psychological review*, vol. 97, no. 3, p. 315, 1990.
- [78] C. Darwin and P. Prodger, *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.
- [79] C. E. Izard, "Basic emotions, relations among emotions, and emotion-cognition relations.," 1992.
- [80] J. Panksepp, *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press, 2004.
- [81] J. E. LeDoux, "Emotion circuits in the brain," *Annual review of neuroscience*, vol. 23, no. 1, pp. 155–184, 2000.
- [82] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American scientist*, vol. 89, no. 4, pp. 344–350, 2001.
- [83] J. R. Averill, "A constructivist view of emotion," in *Theories of emotion*, Elsevier, 1980, pp. 305–339.
- [84] C. Heyes, "Précis of cognitive gadgets: The cultural evolution of thinking," *Behavioral and Brain Sciences*, vol. 42, 2019.
- [85] J. C. Jackson, J. Watts, T. R. Henry, *et al.*, "Emotion semantics show both cultural variation and universal structure," *Science*, vol. 366, no. 6472, pp. 1517–1522, 2019.
- [86] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, "The brain basis of emotion: a meta-analytic review," *The Behavioral and brain sciences*, vol. 35, no. 3, p. 121, 2012.
- [87] W. Lyons, "The philosophy of cognition and emotion," *Handbook of cognition and emotion*, pp. 21–44, 1999.
- [88] R. S. Lazarus, "Emotions and adaptation: Conceptual and empirical relations.," in *Nebraska symposium on motivation*, 1968.
- [89] R. S. Lazarus and R. S. Lazarus, *Emotion and adaptation*. Oxford University Press on Demand, 1991.

- [90] A. Moors, “Automatic constructive appraisal as a candidate cause of emotion,” *Emotion Review*, vol. 2, no. 2, pp. 139–156, 2010.
- [91] E. Harmon-Jones, C. Harmon-Jones, and E. Summerell, “On the importance of both dimensional and discrete models of emotion,” *Behavioral sciences*, vol. 7, no. 4, p. 66, 2017.
- [92] E. L. Rosenberg and P. Ekman, “Coherence between expressive and experiential systems in emotion,” *Cognition & Emotion*, vol. 8, no. 3, pp. 201–229, 1994.
- [93] B. Chakrabarti, E. Bullmore, and S. Baron-Cohen, “Empathizing with basic emotions: common and discrete neural substrates,” *Social neuroscience*, vol. 1, no. 3-4, pp. 364–384, 2006.
- [94] S. Hamann, “Mapping discrete and dimensional emotions onto the brain: controversies and consensus,” *Trends in cognitive sciences*, vol. 16, no. 9, pp. 458–466, 2012.
- [95] W. Sato, S. Hyniewska, K. Minemoto, and S. Yoshikawa, “Facial expressions of basic emotions in Japanese laypeople,” *Frontiers in psychology*, vol. 10, p. 259, 2019.
- [96] J. LeDoux, *The emotional brain: The mysterious underpinnings of emotional life*. Simon and Schuster, 1998.
- [97] C. E. Izard, “Emotion theory and research: Highlights, unanswered questions, and emerging issues,” *Annual review of psychology*, vol. 60, pp. 1–25, 2009.
- [98] K. R. Scherer and A. Moors, “The emotion process: Event appraisal and component differentiation,” *Annual review of psychology*, vol. 70, pp. 719–745, 2019.
- [99] W. G. Parrott, “Components and the definition of emotion,” *Social Science Information*, vol. 46, no. 3, pp. 419–423, 2007.
- [100] R. Giner-Sorolla, “The past thirty years of emotion research: appraisal and beyond,” *Cognition and Emotion*, vol. 33, no. 1, pp. 48–54, 2019.
- [101] P. Kuppens, “Improving theory, measurement, and reality to advance the future of emotion research,” *Cognition and Emotion*, vol. 33, no. 1, pp. 20–23, 2019.
- [102] D. Sander, D. Grandjean, and K. R. Scherer, “A systems approach to appraisal mechanisms in emotion,” *Neural networks*, vol. 18, no. 4, pp. 317–352, 2005.

-
- [103] R. M. Shiffrin and W. Schneider, “Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory.,” *Psychological review*, vol. 84, no. 2, p. 127, 1977.
- [104] H. Leventhal and K. Scherer, “The relationship of emotion to cognition: A functional approach to a semantic controversy,” *Cognition and emotion*, vol. 1, no. 1, pp. 3–28, 1987.
- [105] N. H. Frijda, “Emotions, individual differences and time course: Reflections,” *Cognition and Emotion*, vol. 23, no. 7, pp. 1444–1461, 2009.
- [106] B. Mesquita and M. Boiger, “Emotions in context: A sociodynamic model of emotions,” *Emotion Review*, vol. 6, no. 4, pp. 298–302, 2014.
- [107] M. Bruder, D. Dosmukhambetova, J. Nerb, and A. S. R. Manstead, “Emotional signals in nonverbal interaction: Dyadic facilitation and convergence in expressions, appraisals, and feelings,” *Cognition & emotion*, vol. 26, no. 3, pp. 480–502, 2012.
- [108] T. Imada and P. C. Ellsworth, “Proud Americans and lucky Japanese: Cultural differences in appraisal and corresponding emotion.,” *Emotion*, vol. 11, no. 2, p. 329, 2011.
- [109] P. Wright and J. McCarthy, “Empathy and experience in HCI,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2008, pp. 637–646.
- [110] E. L.-C. Law and P. Van Schaik, “Modelling user experience—An agenda for research and practice,” *Interacting with computers*, vol. 22, no. 5, pp. 313–322, 2010.
- [111] S. Hamann and T. Canli, “Individual differences in emotion processing,” *Current opinion in neurobiology*, vol. 14, no. 2, pp. 233–238, 2004.
- [112] L. E. Williams, J. A. Oler, A. S. Fox, *et al.*, “Fear of the unknown: uncertain anticipation reveals amygdala alterations in childhood anxiety disorders,” *Neuropsychopharmacology*, vol. 40, no. 6, pp. 1428–1435, 2015.
- [113] E. Asutay, A. Genevsky, L. F. Barrett, J. P. Hamilton, P. Slovic, and D. Västfjäll, “Affective calculus: The construction of affect through information integration over time.,” *Emotion*, 2019.
- [114] R. Raghunathan, M. T. Pham, and K. P. Corfman, “Informational properties of anxiety and sadness, and displaced coping,” *Journal of Consumer Research*, vol. 32, no. 4, pp. 596–601, 2006.

- [115] L. K. Hildebrandt, C. McCall, H. G. Engen, and T. Singer, “Cognitive flexibility, heart rate variability, and resilience predict fine-grained regulation of arousal during prolonged threat,” *Psychophysiology*, vol. 53, no. 6, pp. 880–890, 2016.
- [116] P. E. Griffiths, *What emotions really are*. University of Chicago Press, 2008.
- [117] P. M. A. Desmet and P. Hekkert, “Special issue editorial: Design & emotion,” *International Journal of Design*, vol. 3, no. 2, 2009.
- [118] P. Desmet, *Designing emotions*, 2002.
- [119] J. McCarthy and P. Wright, “Technology as experience,” *interactions*, vol. 11, no. 5, pp. 42–43, 2004.
- [120] P. Desmet and P. Hekkert, “Framework of product experience,” *International journal of design*, vol. 1, no. 1, 2007.
- [121] A. G. Ho and K. W. M. G. Siu, “Emotion design, emotional design, emotionalize design: A review on their relationships from a new perspective,” *The Design Journal*, vol. 15, no. 1, pp. 9–32, 2012.
- [122] C. Abras, D. Maloney-Krichmar, J. Preece, *et al.*, “User-centered design,” *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, vol. 37, no. 4, pp. 445–456, 2004.
- [123] J. Mao, B. Horan, H. Forbes, *et al.*, “Application of emotional design to the form redesign of a midwifery training aid,” in *The International Conference on Design and Technology*, 2017, pp. 44–50.
- [124] S. F. Pengnate and R. Sarathy, “An experimental investigation of the influence of website emotional design features on trust in unfamiliar online vendors,” *Computers in Human Behavior*, vol. 67, pp. 49–60, 2017.
- [125] A. Sonderegger and J. Sauer, “The influence of design aesthetics in usability testing: Effects on user performance and perceived usability,” *Applied ergonomics*, vol. 41, no. 3, pp. 403–410, 2010.
- [126] T. Olsson and M. Salo, “Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 2779–2788.
- [127] A. Kultima, “Game design research,” in *Proceedings of the 19th International Academic Mindtrek Conference*, 2015, pp. 18–25.

-
- [128] K. Salen, K. S. Tekinbaş, and E. Zimmerman, *Rules of play: Game design fundamentals*. MIT press, 2004.
- [129] U. Hagen, “Designing for player experience: How professional game developers communicate design visions,” *Journal of Gaming & Virtual Worlds*, vol. 3, no. 3, pp. 259–275, 2011.
- [130] D. Freeman, “Creating emotion in games: The craft and art of emotioneering™,” *Computers in Entertainment (CIE)*, vol. 2, no. 3, p. 15, 2004.
- [131] K. Isbister, *How games move us: Emotion by design*. Mit Press, 2016.
- [132] D. Thue, V. Bulitko, M. Spetch, and T. Romanuik, “Player agency and the relevance of decisions,” in *Joint International Conference on Interactive Digital Storytelling*, 2010, pp. 210–215.
- [133] M. Csikszentmihalyi, “The flow experience and its significance for human psychology,” *Optimal experience: Psychological studies of flow in consciousness*, vol. 2, pp. 15–35, 1988.
- [134] J. Chen, “Flow in games (and everything else),” *Communications of the ACM*, vol. 50, no. 4, pp. 31–34, 2007.
- [135] S. Deterding, “The lens of intrinsic skill atoms: A method for gameful design,” *Human–Computer Interaction*, vol. 30, no. 3-4, pp. 294–335, 2015.
- [136] S. Rigby and R. M. Ryan, *Glued to games: How video games draw us in and hold us spellbound: How video games draw us in and hold us spellbound*. AbC-CLIo, 2011.
- [137] E. L. Deci and R. M. Ryan, “Motivation, personality, and development within embedded social contexts: An overview of self-determination theory,” 2012.
- [138] S. N. Baharom, W. H. Tan, and M. Z. Idris, “Emotional design for games: A framework for player-centric approach in the game design process,” *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 10, pp. 387–398, 2014.
- [139] P. De Byl, “A conceptual affective design framework for the use of emotions in computer game design,” *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 9, no. 3, 2015.
- [140] R. Hunicke, M. LeBlanc, and R. Zubek, “MDA: A formal approach to game design and game research,” in *Proceedings of the AAAI Workshop on Challenges in Game AI*, vol. 4, 2004, p. 1722.

- [141] M. Pichlmair and M. Johansen, “Designing Game Feel. A Survey,” *IEEE Transactions on Games*, 2021.
- [142] E. Hudlicka, “Affective game engines: motivation and requirements,” in *Proceedings of the 4th international conference on foundations of digital games*, 2009, pp. 299–306.
- [143] E. Lux, M. T. P. Adam, V. Dorner, S. Helming, M. T. Knierim, and C. Weinhardt, “Live biofeedback as a user interface design element: A review of the literature,” *Communications of the Association for Information Systems*, vol. 43, no. 1, p. 18, 2018.
- [144] Y. Y. Ng and C. W. Khong, “A review of affective user-centered design for video games,” in *2014 3rd international conference on user science and engineering (i-user)*, 2014, pp. 79–84.
- [145] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, “Human emotion recognition: Review of sensors and methods,” *Sensors*, vol. 20, no. 3, p. 592, 2020.
- [146] S. Hamdy and D. King, “Affect and believability in game characters—a review of the use of affective computing in games,” in *Proceedings of the 18th Annual Conference on Simulation and AI in Computer Games. EUROSIS*, 2017.
- [147] Y. Wang, W. Song, W. Tao, *et al.*, “A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances,” *Information Fusion*, 2022.
- [148] K. M. Gilleade and A. Dix, “Using frustration in the design of adaptive videogames,” in *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 2004, pp. 228–232.
- [149] A. Picardi, P. Burelli, and G. N. Yannakakis, “Modelling virtual camera behaviour through player gaze,” in *Proceedings of the 6th international conference on foundations of digital games*, 2011, pp. 107–114.
- [150] B. Perron, *Silent hill: The terror engine*. University of Michigan Press, 2012.
- [151] L. Anolli, F. Mantovani, L. Confalonieri, A. Ascolese, and L. Peveri, “Emotions in serious games: From experience to assessment,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 5, no. 2010, 2010.
- [152] H. Al Osman, H. Dong, and A. El Saddik, “Ubiquitous biofeedback serious game for stress management,” *IEEE Access*, vol. 4, pp. 1274–1286, 2016.

-
- [153] P. Jerčić, P. J. Astor, M. Adam, *et al.*, “A serious game using physiological interfaces for emotion regulation training in the context of financial decision-making,” in *20th European Conference on Information Systems (ECIS 2012)*, Barcelona, 2012, pp. 1–14.
- [154] O. Hilborn, H. Cederholm, J. Eriksson, and C. Lindley, “A biofeedback game for training arousal regulation during a stressful task: The space investor,” in *International Conference on Human-Computer Interaction*, 2013, pp. 403–410.
- [155] H. Cederholm, O. Hilborn, C. Lindley, C. C. Sennersten, and J. Eriksson, “The Aiming Game: Using a Game with Biofeedback for Training in Emotion Regulation.,” in *DiGRA Conference*, 2011.
- [156] G. N. Yannakakis, H. P. Martinez, and A. Jhala, “Towards affective camera control in games,” *User Modeling and User-Adapted Interaction*, vol. 20, no. 4, pp. 313–340, 2010.
- [157] Y. A. Sekhavat, M. J. Sisi, and S. Roohi, “Affective interaction: Using emotions as a user interface in games,” *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5225–5253, 2021.
- [158] G. N. Yannakakis and J. Togelius, “Experience-driven procedural content generation,” *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 147–161, 2011.
- [159] D. Bersak, G. McDarby, N. Augenblick, *et al.*, “Intelligent biofeedback using an immersive competitive environment,” in *Paper at the designing ubiquitous computing games workshop at UbiComp*, 2001, pp. 1–6.
- [160] R. Games and Rockstar Games, *Red Dead Redemption 2*, [Game], 2018.
- [161] C. D. Projekt and CD Projekt, *Cyberpunk 77*, [Game], 2020.
- [162] M. Egger, M. Ley, and S. Hanke, “Emotion recognition from physiological signal analysis: a review,” *Electronic Notes in Theoretical Computer Science*, vol. 343, pp. 35–55, 2019.
- [163] J. M. Kivikangas, G. Chanel, B. Cowley, *et al.*, “A review of the use of psychophysiological methods in game research,” *journal of gaming & virtual worlds*, vol. 3, no. 3, pp. 181–199, 2011.

- [164] E. Harmon-Jones, “On motivational influences, moving beyond valence, and integrating dimensional and discrete views of emotion,” *Cognition and Emotion*, vol. 33, no. 1, pp. 101–108, 2019.
- [165] W. Mellouk and W. Handouzi, “Facial emotion recognition using deep learning: review and insights,” *Procedia Computer Science*, vol. 175, pp. 689–694, 2020.
- [166] P. M. Blom, S. Bakkes, C. Tan, *et al.*, “Towards personalised gaming via facial expression recognition,” in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 10, 2014.
- [167] B. Allaert, J. Mennesson, and I. M. Bilasco, “EmoGame: towards a self-rewarding methodology for capturing children faces in an engaging context,” in *International Workshop on Human Behavior Understanding*, Springer, 2016, pp. 3–14.
- [168] K. Boehner, R. DePaula, P. Dourish, and P. Sengers, “How emotion is made and measured,” *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 275–291, 2007.
- [169] J. T. Hancock, K. Gee, K. Ciaccio, and J. M.-H. Lin, “I’m sad you’re sad: emotional contagion in CMC,” in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 2008, pp. 295–298.
- [170] S. Harrer, *Games and Bereavement: How Video Games Represent Attachment, Loss and Grief*. transcript Verlag, 2018.
- [171] R. Beale and C. Creed, “Affective interaction: How emotional agents affect users,” *International journal of human-computer studies*, vol. 67, no. 9, pp. 755–776, 2009.
- [172] Y. Ma, K. L. Nguyen, F. Z. Xing, and E. Cambria, “A survey on empathetic dialogue systems,” *Information Fusion*, vol. 64, pp. 50–70, 2020.
- [173] A. Popescu, J. Broekens, and M. Van Someren, “Gamygdala: An emotion engine for games,” *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 32–44, 2013.
- [174] B. Brewster, “Portfolio,” *The Yale Literary Magazine*, pp. 201–205, Feb. 1882.
- [175] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group*, “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement,” *Annals of internal medicine*, vol. 151, no. 4, pp. 264–269, 2009.
- [176] OSF, *Affective Theories, Methodology, and Evaluation of Emotion-Adaptive Games: A Systematic Review*, 2022. [Online]. Available: <https://osf.io/qep2u>.

-
- [177] C. O. Fritz, P. E. Morris, and J. J. Richler, “Effect size estimates: current use, calculations, and interpretation,” *Journal of experimental psychology: General*, vol. 141, no. 1, p. 2, 2012.
- [178] T. Baguley, “Understanding statistical power in the context of applied research,” *Applied ergonomics*, vol. 35, no. 2, pp. 73–80, 2004.
- [179] J. P. T. Higgins, J. Thomas, J. Chandler, *et al.*, *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons, 2019.
- [180] W. A. IJsselsteijn, Y. A. W. De Kort, and K. Poels, “The game experience questionnaire,” 2013.
- [181] D. Markland and L. Hardy, “On the factorial and construct validity of the Intrinsic Motivation Inventory: Conceptual and operational concerns,” *Research quarterly for exercise and sport*, vol. 68, no. 1, pp. 20–32, 1997.
- [182] H.-Y. Sung, G.-J. Hwang, and Y.-F. Yen, “Development of a contextual decision-making game for improving students’ learning performance in a health education course,” *Computers & Education*, vol. 82, pp. 179–190, 2015.
- [183] C. Jennett, A. L. Cox, P. Cairns, *et al.*, “Measuring and defining the experience of immersion in games,” *International journal of human-computer studies*, vol. 66, no. 9, pp. 641–661, 2008.
- [184] V. Vanden Abeele, L. E. Nacke, E. D. Mekler, and D. Johnson, “Design and preliminary validation of the player experience inventory,” in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 2016, pp. 335–341.
- [185] I. E. Vermeulen, C. Roth, P. Vorderer, and C. Klimmt, “Measuring user responses to interactive stories: Towards a standardized assessment tool,” in *Joint International Conference on Interactive Digital Storytelling*, 2010, pp. 38–43.
- [186] P. Gomes, A. Paiva, C. Martinho, and A. Jhala, “Metrics for character believability in interactive narrative,” in *International conference on interactive digital storytelling*, 2013, pp. 223–228.
- [187] M. T. Akbar, M. N. Ilmi, I. V. Rumayar, J. Moniaga, T.-K. Chen, and A. Chowanda, “Enhancing game experience with facial expression recognition as dynamic balancing,” *Procedia Computer Science*, vol. 157, pp. 388–395, 2019.

- [188] T. Alves, S. Gama, and F. S. Melo, "Flow adaptation in serious games for health," in *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*, 2018, pp. 1–8.
- [189] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing," *Journal of experimental social psychology*, vol. 45, no. 4, pp. 624–634, 2009.
- [190] A. N. T. Andrew and A. Chowanda, "DYNAMIC DIFFICULTY ADJUSTMENT WITH FACIAL EXPRESSION RECOGNITION FOR IMPROVING PLAYER SATISFACTION IN A SURVIVAL HORROR GAME," *ICIC Express Letters*, vol. 14, 2020.
- [191] P. M. Blom, S. Bakkes, and P. Spronck, "Modeling and adjusting in-game difficulty based on facial expression analysis," *Entertainment Computing*, vol. 31, p. 100307, 2019.
- [192] B. Bontchev and D. Vassileva, "Affect-based adaptation of an applied video game for educational purposes," *Interactive Technology and Smart Education*, 2017.
- [193] B. Bontchev and O. Georgieva, "Playing style recognition through an adaptive video game," *Computers in Human Behavior*, vol. 82, pp. 136–147, 2018.
- [194] A. Darzi, S. M. McCrea, D. Novak, *et al.*, "User experience with dynamic difficulty adjustment methods for an affective exergame: Comparative laboratory-based study," *JMIR Serious Games*, vol. 9, no. 2, e25771, 2021.
- [195] K. C. Ewing, S. H. Fairclough, and K. Gilleade, "Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a Bio-cybernetic loop," *Frontiers in Human Neuroscience*, vol. 10, no. MAY2016, p. 223, 2016. DOI: 10.3389/fnhum.2016.00223.
- [196] G. Matthews, D. M. Jones, and A. G. Chamberlain, "Refining the measurement of mood: The UWIST mood adjective checklist," *British journal of psychology*, vol. 81, no. 1, pp. 17–42, 1990.
- [197] J. Frommel, F. Fischbach, K. Rogers, and M. Weber, "Emotion-based dynamic difficulty adjustment using parameterized difficulty and self-reports of emotion," in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 2018, pp. 163–171.

-
- [198] E. McAuley, T. Duncan, and V. V. Tammen, “Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis,” *Research quarterly for exercise and sport*, vol. 60, no. 1, pp. 48–58, 1989.
- [199] S. P. Hernandez, V. Bulitko, and M. Spetch, “Keeping the player on an emotional trajectory in interactive storytelling,” in *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2015.
- [200] M. L. Ibáñez, M. Miranda, N. Alvarez, and F. Peinado, “Using gestural emotions recognised through a neural network as input for an adaptive music system in virtual reality,” *Entertainment Computing*, vol. 38, p. 100 404, 2021.
- [201] M. Slater, M. Usoh, and A. Steed, “Depth of presence in virtual environments,” *Presence: Teleoperators & Virtual Environments*, vol. 3, no. 2, pp. 130–144, 1994.
- [202] J. Jalbert and S. Rank, “Exit 53: physiological data for improving non-player character interaction,” in *International Conference on Interactive Digital Storytelling*, 2016, pp. 25–36.
- [203] C. Lara-Alvarez, H. Mitre-Hernandez, J. J. Flores, and H. Pérez-Espinosa, “Induction of emotional states in educational video games through a fuzzy control system,” *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 66–77, 2018.
- [204] C. Liu, P. Agrawal, N. Sarkar, and S. Chen, “Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback,” *International Journal of Human-Computer Interaction*, vol. 25, no. 6, pp. 506–529, 2009.
- [205] J. V. Moniaga, A. Chowanda, A. Prima, M. D. T. Rizqi, *et al.*, “Facial expression recognition as dynamic game balancing system,” *Procedia Computer Science*, vol. 135, pp. 361–368, 2018.
- [206] F. Negini, R. L. Mandryk, and K. G. Stanley, “Using affective state to adapt characters, NPCs, and the environment in a first-person shooter game,” in *2014 IEEE Games Media Entertainment*, 2014, pp. 1–8.
- [207] P. A. Nogueira, V. Torres, R. Rodrigues, E. Oliveira, and L. E. Nacke, “Vanishing scares: biofeedback modulation of affective player experiences in a procedural horror game,” *Journal on Multimodal User Interfaces*, vol. 10, no. 1, pp. 31–62, 2016.

- [208] A. Parnandi, B. Ahmed, E. Shipp, and R. Gutierrez-Osuna, "Chill-Out: Relaxation training through respiratory biofeedback in a mobile casual game," in *International Conference on Mobile Computing, Applications, and Services*, 2013, pp. 252–260.
- [209] A. Parnandi and R. Gutierrez-Osuna, "Physiological modalities for relaxation skill transfer in biofeedback games," *IEEE journal of biomedical and health informatics*, vol. 21, no. 2, pp. 361–371, 2015.
- [210] C. Rodriguez-Guerrero, K. Knaepen, J. C. Fraile-Marinero, J. Perez-Turiel, V. Gonzalez-de-Garibay, and D. Lefeber, "Improving challenge/skill ratio in a multimodal interface by simultaneously adapting game difficulty and haptic assistance through psychophysiological and performance feedback," *Frontiers in neuroscience*, vol. 11, p. 242, 2017.
- [211] M. P. C. Rosa, E. A. d. Santos, I. L. R. de Moraes, M. M. Sarmet, C. D. Castanho, R. P. Jacobi, *et al.*, "Dynamic Difficulty Adjustment Using Performance and Affective Data in a Platform Game," in *International Conference on Human-Computer Interaction*, 2021, pp. 367–386.
- [212] J. Salah, Y. Abdelrahman, A. Dakrouni, and S. Abdennadher, "Judged by the Cover: Investigating the Effect of Adaptive Game Interface on the Learning Experience," in *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, 2018, pp. 215–225.
- [213] A. Stein, Y. Yotam, R. Puzis, G. Shani, and M. Taieb-Maimon, "EEG-triggered dynamic difficulty adjustment for multiplayer games," *Entertainment computing*, vol. 25, pp. 14–25, 2018.
- [214] A. N. Tjokrosetio and A. Chowanda, "Character believability enhancement using facial expression recognition to improve the players immersive experience," *ICIC Express Letters*, vol. 15, pp. 1235–1242, 2021.
- [215] V. Vachiratamporn, K. Moriyama, K.-i. Fukui, and M. Numao, "An implementation of affective adaptation in survival horror games," in *2014 IEEE conference on computational intelligence and games*, 2014, pp. 1–8.
- [216] A. Denisova, A. I. Nordin, and P. Cairns, "The convergence of player experience questionnaires," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 2016, pp. 33–37.

- [217] D. Johnson, M. J. Gardner, and R. Perry, “Validation of two game experience scales: the player experience of need satisfaction (PENS) and game experience questionnaire (GEQ),” *International Journal of Human-Computer Studies*, vol. 118, pp. 38–46, 2018.
- [218] L. Michailidis, E. Balaguer-Ballester, and X. He, “Flow and immersion in video games: The aftermath of a conceptual challenge,” *Frontiers in psychology*, vol. 9, p. 1682, 2018.
- [219] S. E. Maxwell, K. Kelley, and J. R. Rausch, “Sample size planning for statistical power and accuracy in parameter estimation,” *Annu. Rev. Psychol.*, vol. 59, pp. 537–563, 2008.
- [220] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, *Introduction to meta-analysis*. John Wiley & Sons, 2021.
- [221] J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” *Information Fusion*, vol. 59, pp. 103–126, 2020.
- [222] M. Gendron, D. Roberson, J. M. van der Vyver, and L. F. Barrett, “Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture,” *Emotion*, vol. 14, no. 2, p. 251, 2014.
- [223] K. Jalife, C. Harteveld, and C. Holmgård, “From Flow to Fuse: A Cognitive Perspective,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CHI PLAY, pp. 1–30, 2021.
- [224] C. J. Fong, D. J. Zaleski, and J. K. Leach, “The challenge–skill balance and antecedents of flow: A meta-analytic investigation,” *The Journal of Positive Psychology*, vol. 10, no. 5, pp. 425–446, 2015.
- [225] R. W. Picard, “What does it mean for a computer to “have” emotions,” *Emotions in humans and artifacts*, pp. 213–235, 2003.
- [226] L. Li and J. Campbell, “Emotion modeling and interaction of NPCS in virtual simulation and games,” *International Journal of Virtual Reality*, vol. 9, no. 4, pp. 1–6, 2010.

- [227] M. Causse, B. Pavard, J.-M. Senard, J.-F. Démonet, and J. Pastor, “Positive and negative emotion induction through avatars and its impact on reasoning performance: cardiovascular and pupillary correlates,” *Studia Psychologica*, vol. 54, pp. 37, 2012.
- [228] P. Lankoski, “Computer games and emotions,” in *The philosophy of computer games*, Springer, 2012, pp. 39–55.
- [229] Y. Ng, C. W. Khong, and R. J. Nathan, “Evaluating affective user-centered design of video games using qualitative methods,” *International Journal of Computer Games Technology*, vol. 2018, 2018.
- [230] M. K. Miller and R. L. Mandryk, “Differentiating in-game frustration from at-game frustration using touch pressure,” in *Proceedings of the 2016 ACM international conference on interactive surfaces and spaces*, 2016, pp. 225–234.
- [231] N. Shaker, S. Asteriadis, G. N. Yannakakis, and K. Karpouzis, “Fusing visual and behavioral cues for modeling user experience in games,” *IEEE Transactions on cybernetics*, vol. 43, no. 6, pp. 1519–1531, 2013.
- [232] G. F. Tondello, “Dynamic personalization of gameful interactive systems,” *University of Waterloo, Waterloo, Canada Phd thesis*, 2019.
- [233] J. S. Dumas, J. S. Dumas, and J. Redish, *A practical guide to usability testing*. Intellect books, 1999.
- [234] G. Salvendy, *Handbook of human factors and ergonomics*. John Wiley & Sons, 2012.
- [235] A. Denisova, J. A. Bopp, T. D. Nguyen, and E. D. Mekler, “Whatever the emotional experience, it’s up to them: Insights from designers of emotionally impactful games,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2021. DOI: 10.1145/3411764.3445286.
- [236] J. A. Bopp, E. D. Mekler, and K. Opwis, “Negative emotion, positive experience? Emotionally moving moments in digital games,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 2996–3006.
- [237] A. Tyack and E. D. Mekler, “Off-Peak: An Examination of Ordinary Player Experience,” 2021.
- [238] Y. Li, W. Ma, Q. Kang, *et al.*, “Night or darkness, which intensifies the feeling of fear?” *International Journal of Psychophysiology*, vol. 97, no. 1, pp. 46–57, 2015.

- [239] S. Goodman, *Sonic warfare: Sound, affect, and the ecology of fear*. mit Press, 2012.
- [240] M. A. Urista, Q. Dong, K. D. Day, *et al.*, “Explaining why young adults use MySpace and Facebook through uses and gratifications theory,” *Human Communication*, vol. 12, no. 2, pp. 215–229, 2009.
- [241] V. Zammito, “Games User Research as part of the development process in the game industry: Challenges and best practices,” *Games User Research, Anders Drachen, Pejman Mirza-Babaei, and Lennart E. Nacke (Eds.). Oxford University Press, Oxford, UK*, pp. 15–30, 2018.
- [242] A. Drachen, P. Mirza-Babaei, and L. E. Nacke, *Games user research*. Oxford University Press, 2018.
- [243] E. Schubert, “Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space,” *Australian Journal of Psychology*, vol. 51, no. 3, pp. 154–165, 1999.
- [244] A. M. Ruef and R. W. Levenson, “Continuous measurement of emotion,” *Handbook of emotion elicitation and assessment*, pp. 286–297, 2007.
- [245] W. Hollway and T. Jefferson, *Doing qualitative research differently: Free association, narrative and the interview method*. Sage, 2000.
- [246] A. S. Cowen and D. Keltner, “Self-report captures 27 distinct categories of emotion bridged by continuous gradients,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 38, E7900–E7909, 2017.
- [247] J. T. Alexander, J. Sear, and A. Oikonomou, “An investigation of the effects of game difficulty on player enjoyment,” *Entertainment Computing*, vol. 4, no. 1, pp. 53–62, 2013. DOI: 10.1016/j.entcom.2012.09.001.
- [248] A. Dekker and E. Champion, “Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback,” in *DiGRA ’07-Proceedings of the 2007 DiGRA International Conference: Situated Play*, 2007, pp. 550–558.
- [249] T. Christy and L. I. Kuncheva, “Technological advancements in affective gaming: A historical survey,” *GSTF Journal on Computing*, vol. 3, no. 4, pp. 32–41, 2014.
- [250] T. Tijs, D. Brokken, and W. IJsselsteijn, “Creating an emotionally adaptive game,” in *International Conference on Entertainment Computing*, 2008, pp. 122–133.

- [251] J. Benedek and R. Hazlett, "Incorporating facial emg emotion measures as feedback in the software design process," *Proc. Human Computer Interaction Consortium*, 2005.
- [252] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological psychology*, vol. 84, no. 3, pp. 394–421, 2010.
- [253] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience," *IEEE Transactions on Computational Intelligence and AI in games*, vol. 4, no. 3, pp. 199–212, 2012.
- [254] L. F. Barrett and A. B. Satpute, "Historical pitfalls and new directions in the neuroscience of emotion," *Neuroscience letters*, vol. 693, pp. 9–18, 2019.
- [255] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 211–223, 2011.
- [256] J. Z. Lim, J. Mountstephens, and J. Teo, "Emotion recognition using eye-tracking: taxonomy, review and current challenges," *Sensors*, vol. 20, no. 8, p. 2384, 2020.
- [257] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk, "Biofeedback game design: using direct and indirect physiological control to enhance game interaction," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 103–112.
- [258] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghamurst, "Exploring natural eye-gaze-based interaction for immersive virtual reality," in *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 2017, pp. 36–39.
- [259] J. L. Soler-Dominguez, J. D. Camba, M. Contero, and M. Alcañiz, "A proposal for the selection of eye-tracking metrics for the implementation of adaptive gameplay in virtual reality based games," in *International Conference on Virtual, Augmented and Mixed Reality*, 2017, pp. 369–380.
- [260] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What does touch tell us about emotions in touchscreen-based gameplay?" *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 19, no. 4, pp. 1–30, 2012.

- [261] J. Sykes and S. Brown, “Affective gaming: measuring emotion through the gamepad,” in *CHI’03 extended abstracts on Human factors in computing systems*, 2003, pp. 732–733.
- [262] J. Frommel, C. Schrader, and M. Weber, “Towards emotion-based adaptive games: Emotion recognition via input and performance features,” in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 2018, pp. 173–185.
- [263] M. Kim and Y. Y. Doh, “Computational modeling of players’ emotional response patterns to the story events of video games,” *IEEE Transactions on Affective Computing*, vol. 8, no. 2, pp. 216–227, 2017.
- [264] A. Canossa, *Play-Persona: Modeling Player Behavior in Computer Games*. Danmarks Designskole, 2009.
- [265] N. Shaker, G. Yannakakis, and J. Togelius, “Towards automatic personalized content generation for platform games,” in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 5, 2010.
- [266] R. Reizenzein, E. Hudlicka, M. Dastani, *et al.*, “Computational modeling of emotion: Toward improving the inter-and intradisciplinary exchange,” *IEEE Transactions on Affective Computing*, vol. 4, no. 3, pp. 246–266, 2013.
- [267] L. Caroux, K. Isbister, L. Le Bigot, and N. Vibert, “Player–video game interaction: A systematic review of current concepts,” *Computers in Human Behavior*, vol. 48, pp. 366–381, 2015.
- [268] E. D. Mekler, J. A. Bopp, A. N. Tuch, and K. Opwis, “A systematic review of quantitative studies on the enjoyment of digital entertainment games,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2014, pp. 927–936.
- [269] M. Zohaib, “Dynamic difficulty adjustment (DDA) in computer games: A review,” *Advances in Human-Computer Interaction*, vol. 2018, 2018. DOI: 10.1155/2018/5681652.
- [270] G. Husain, W. F. Thompson, and E. G. Schellenberg, “Effects of musical tempo and mode on arousal, mood, and spatial abilities,” *Music perception*, vol. 20, no. 2, pp. 151–171, 2002.

- [271] S. R. Livingstone and A. R. Brown, “Dynamic Response: Real-Time Adaptation for Music Emotion,” in *Proceedings of the Second Australasian Conference on Interactive Entertainment*, ser. IE '05, Sydney, AUS: Creativity & Cognition Studios Press, 2005, pp. 105–111.
- [272] K. Rogers, M. Jörg, and M. Weber, “Effects of background music on risk-taking and general player experience,” in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 2019, pp. 213–224.
- [273] D. Gundry and S. Deterding, “Validity threats in quantitative data collection with games: A narrative survey,” *Simulation & Gaming*, vol. 50, no. 3, pp. 302–328, 2019.
- [274] R. P. McMahan, E. D. Ragan, A. Leal, R. J. Beaton, and D. A. Bowman, “Considerations for the use of commercial video games in controlled experiments,” *Entertainment Computing*, vol. 2, no. 1, pp. 3–9, 2011.
- [275] A. Tyack, P. Wyeth, and M. Klarkowski, “Video game selection procedures for experimental research,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–9.
- [276] J. A. Russell, A. Weiss, and G. A. Mendelsohn, “Affect grid: a single-item scale of pleasure and arousal.,” *Journal of personality and social psychology*, vol. 57, no. 3, p. 493, 1989.
- [277] C. J. Ferguson, “An effect size primer: a guide for clinicians and researchers.,” 2016.
- [278] A. P. Dijksterhuis and P. K. Smith, “Affective habituation: subliminal exposure to extreme stimuli decreases their extremity.,” *Emotion*, vol. 2, no. 3, p. 203, 2002.
- [279] A. Aljanaki, F. Wiering, and R. C. Veltkamp, “Studying emotion induced by music through a crowdsourcing game,” *Information Processing & Management*, vol. 52, no. 1, pp. 115–128, 2016.
- [280] R. Vicencio-Moreira, R. L. Mandryk, and C. Gutwin, “Now you can compete with anyone: Balancing players of different skill levels in a first-person shooter game,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 2255–2264.
- [281] H. Martineau, *How to observe morals and manners*. Transaction Publishers, 1838, vol. 1.

- [282] J. Hanich, V. Wagner, M. Shah, T. Jacobsen, and W. Menninghaus, "Why we like to watch sad films. The pleasure of being moved in aesthetic experiences.," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 8, no. 2, p. 130, 2014.
- [283] L. Reinecke, "Games and recovery: The use of video and computer games to recuperate from stress and strain," *Journal of Media Psychology*, vol. 21, no. 3, pp. 126–142, 2009.
- [284] C. J. Ferguson and C. K. Olson, "Friends, fun, frustration and fantasy: Child motivations for video game play," *Motivation and Emotion*, vol. 37, no. 1, pp. 154–164, 2013.
- [285] S. K. Whitbourne, S. Ellenberg, and K. Akimoto, "Reasons for playing casual video games and perceived benefits among adults 18 to 80 years old," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 12, pp. 892–897, 2013.
- [286] A. Waszkiewicz and M. Bakun, "Towards the aesthetics of cozy video games," *Journal of Gaming & Virtual Worlds*, vol. 12, no. 3, pp. 225–240, 2020.
- [287] European Centre for Disease Prevention and Control, *COVID-19 situation update worldwide as of 1 December 2020*, 2020. [Online]. Available: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>.
- [288] R. P. Rajkumar, "COVID-19 and mental health: A review of the existing literature," *Asian journal of psychiatry*, vol. 52, p. 102066, 2020. DOI: 10.1016/j.ajp.2020.102066.
- [289] N. Salari, A. Hosseinian-Far, R. Jalali, *et al.*, "Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis," *Globalization and Health*, vol. 16, no. 1, p. 57, 2020. DOI: 10.1186/s12992-020-00589-w.
- [290] J. Xiong, O. Lipsitz, F. Nasri, *et al.*, "Impact of COVID-19 pandemic on mental health in the general population: A systematic review," *Journal of affective disorders*, vol. 277, pp. 55–64, 2020. DOI: 10.1016/j.jad.2020.08.001.
- [291] R. C. Kessler, "The effects of stressful life events on depression," *Annual review of psychology*, vol. 48, no. 1, pp. 191–214, 1997.
- [292] C. Tennant, "Life events, stress and depression: a review of recent findings," *Australian & New Zealand Journal of Psychiatry*, vol. 36, no. 2, pp. 173–182, 2002.

BIBLIOGRAPHY

- [293] R. H. Rahe, "Life change events and mental illness: an overview," *Journal of Human Stress*, vol. 5, no. 3, pp. 2–10, 1979.
- [294] C. Faravelli, C. Lo Sauro, L. Lelli, *et al.*, "The role of life events and HPA axis in anxiety disorders: a review," *Current pharmaceutical design*, vol. 18, no. 35, p. 5663, 2012.
- [295] E. S. Paykel, "Life events and affective disorders," *Acta Psychiatrica Scandinavica*, vol. 108, pp. 61–66, 2003.
- [296] E. Suh, E. Diener, and F. Fujita, "Events and subjective well-being: Only recent events matter," *Journal of personality and social psychology*, vol. 70, no. 5, p. 1091, 1996.
- [297] H. H. Schiffrin and S. K. Nelson, "Stressed and happy? Investigating the relationship between happiness and perceived stress," *Journal of Happiness Studies*, vol. 11, no. 1, pp. 33–39, 2010.
- [298] T. Esch, G. B. Stefano, G. L. Fricchione, and H. Benson, "The role of stress in neurodegenerative diseases and mental disorders," *Neuroendocrinology letters*, vol. 23, no. 3, pp. 199–208, 2002.
- [299] R. Contrada and A. Baum, *The handbook of stress science: Biology, psychology, and health*. Springer Publishing Company, 2010.
- [300] M. Croissant, H. Glaesmer, T. Klucken, *et al.*, "Endocannabinoid concentrations in hair and mental health of unaccompanied refugee minors," *Psychoneuroendocrinology*, vol. 116, p. 104683, 2020.
- [301] R. S. Lazarus, "The psychology of stress and coping," *Issues in Mental Health Nursing*, vol. 7, no. 1-4, pp. 399–418, 1985.
- [302] J. A. Gray, *The psychology of fear and stress*. CUP Archive, 1987, vol. 5.
- [303] C. V. Russoniello, K. O'Brien, and J. M. Parks, "The effectiveness of casual video games in improving mood and decreasing stress," *Journal of CyberTherapy & Rehabilitation*, vol. 2, no. 1, pp. 53–66, 2009.
- [304] E. A. Holmes, E. L. James, T. Coode-Bate, and C. Deeprose, "Can playing the computer game "Tetris" reduce the build-up of flashbacks for trauma? A proposal from cognitive science," *PloS one*, vol. 4, no. 1, e4153, 2009.

- [305] E. A. Holmes, E. L. James, E. J. Kilford, and C. Deepröse, “Key steps in developing a cognitive vaccine against traumatic flashbacks: Visuospatial Tetris versus verbal Pub Quiz,” *PloS one*, vol. 5, no. 11, e13706, 2010.
- [306] L. Iyadurai, S. E. Blackwell, R. Meiser-Stedman, *et al.*, “Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: a proof-of-concept randomized controlled trial,” *Molecular psychiatry*, vol. 23, no. 3, pp. 674–682, 2018.
- [307] L. Reinecke, J. Klatt, and N. C. Krämer, “Entertaining media use and the satisfaction of recovery needs: Recovery outcomes associated with the use of interactive and noninteractive entertaining media,” *Media Psychology*, vol. 14, no. 2, pp. 192–215, 2011.
- [308] C. Romero-Blanco, J. Rodríguez-Almagro, M. D. Onieva-Zafra, M. L. Parra-Fernández, M. D. C. Prado-Laguna, and A. Hernández-Martínez, “Physical activity and sedentary lifestyle in university students: Changes during confinement due to the Covid-19 pandemic,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, p. 6567, 2020.
- [309] F. Gallè, E. A. Sabella, S. Ferracuti, *et al.*, “Sedentary behaviors and physical activity of Italian undergraduate students during lockdown at the time of CoViD 19 pandemic,” *International journal of environmental research and public health*, vol. 17, no. 17, p. 6171, 2020.
- [310] J. Meyer, C. McDowell, J. Lansing, *et al.*, “Changes in physical activity and sedentary behavior in response to COVID-19 and their associations with mental health in 3052 US adults,” *International journal of environmental research and public health*, vol. 17, no. 18, 2020.
- [311] S. Chopra, P. Ranjan, V. Singh, *et al.*, “Impact of COVID-19 on lifestyle-related behaviours—a cross-sectional audit of responses from nine hundred and ninety-five participants from India,” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 6, pp. 2021–2030, 2020.
- [312] F. Qin, Y. Song, G. P. Nassis, *et al.*, “Physical activity, screen time, and emotional well-being during the 2019 novel coronavirus outbreak in China,” *International journal of environmental research and public health*, vol. 17, no. 14, p. 5170, 2020.

BIBLIOGRAPHY

- [313] M. Xiang, Z. Zhang, and K. Kuwahara, “Impact of COVID-19 pandemic on children and adolescents’ lifestyle behavior larger than expected,” *Progress in Cardiovascular Diseases*, 2020.
- [314] OSF, *Has there been an increase in the demand for relaxation games during the COVID-19 pandemic?* 2020. [Online]. Available: <https://osf.io/7j4sz>.
- [315] Steamworks Development, *Steam - 2018 Year in Review*, 2019. [Online]. Available: <https://steamcommunity.com/groups/steamworks/announcements/detail/1697194621363928453>.
- [316] Steamworks Development, *Steam - 2019 Year in Review*, 2020. [Online]. Available: <https://steamcommunity.com/groups/steamworks/announcements/detail/1697229969000435735>.
- [317] W. H. Organization *et al.*, “Mental health and psychosocial considerations during the COVID-19 outbreak, 18 March 2020,” World Health Organization, Tech. Rep., 2020.
- [318] D. E. Ho, K. Imai, G. King, and E. A. Stuart, “Matching as nonparametric pre-processing for reducing model dependence in parametric causal inference,” *Political analysis*, vol. 15, no. 3, pp. 199–236, 2007.
- [319] X. S. Gu and P. R. Rosenbaum, “Comparison of multivariate matching methods: Structures, distances, and algorithms,” *Journal of Computational and Graphical Statistics*, vol. 2, no. 4, pp. 405–420, 1993.
- [320] E. A. Stuart and D. B. Rubin, “Best practices in quasi-experimental designs,” *Best practices in quantitative methods*, pp. 155–176, 2008.
- [321] E. A. Stuart, G. King, K. Imai, and D. Ho, “MatchIt: nonparametric preprocessing for parametric causal inference,” *Journal of statistical software*, 2011.
- [322] D. B. Rubin, “Matching to remove bias in observational studies,” *Biometrics*, pp. 159–183, 1973.
- [323] F. Wilcoxon, “Individual Comparisons by Ranking Methods,” *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945. DOI: 10.2307/3001968.
- [324] R. Rosenthal, “Effect sizes: Pearson’s correlation, its display via the BESD, and alternative indices,” 1991.

-
- [325] J. Cohen, *Statistical power analysis for the behavioral sciences*. Academic press, 2013.
- [326] D. Quade, “Rank analysis of covariance,” *Journal of the American Statistical Association*, vol. 62, no. 320, pp. 1187–1200, 1967.
- [327] A. Sultana, S. Tasnim, S. Bhattacharya, M. M. Hossain, and N. Purohit, “Digital screen time during COVID-19 pandemic: A public health concern,” 2020.
- [328] L. J. Lee Smith, M. Trott, A. Yakkundi, *et al.*, “The association between screen time and mental health during COVID-19: A cross sectional study,” *Psychiatry research*, vol. 292, p. 113 333, 2020.
- [329] E. Cvejic, E. C. Lynar, Y. M. Chung, and U. Vollmer-Conna, “One size does not fit all: Individual differences in cardiac autonomic and subjective responses to brief relaxation activities,” *International journal of cardiology*, vol. 223, pp. 265–267, 2016.
- [330] A. A. Augustine and S. H. Hemenover, “On the relative effectiveness of affect regulation strategies: A meta-analysis,” *Cognition and Emotion*, vol. 23, no. 6, pp. 1181–1220, 2009.
- [331] A. Aldao, S. Nolen-Hoeksema, and S. Schweizer, “Emotion-regulation strategies across psychopathology: A meta-analytic review,” *Clinical psychology review*, vol. 30, no. 2, pp. 217–237, 2010.
- [332] Y. Hasan, L. Bègue, and B. J. Bushman, “Violent video games stress people out and make them more aggressive,” *Aggressive behavior*, vol. 39, no. 1, pp. 64–70, 2013.
- [333] B. Baldaro, G. Tuozi, M. Codispoti, *et al.*, “Aggressive and non-violent videogames: short-term psychological and cardiovascular effects on habitual players,” *Stress and Health: Journal of the International Society for the Investigation of Stress*, vol. 20, no. 4, pp. 203–208, 2004.
- [334] E. Collins and A. L. Cox, “Switch on to games: Can digital games aid post-work recovery?” *International Journal of Human-Computer Studies*, vol. 72, no. 8-9, pp. 654–662, 2014.
- [335] Y. J. Halbrook, A. T. O’Donnell, and R. M. Msetfi, “When and how video games can be good: A review of the positive effects of video games on well-being,” *Perspectives on Psychological Science*, vol. 14, no. 6, pp. 1096–1104, 2019.

- [336] T. Kikusui, J. T. Winslow, and Y. Mori, “Social buffering: relief from stress and anxiety,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1476, pp. 2215–2228, 2006.
- [337] M. Yao, Y. Zhou, J. Li, and X. Gao, “Violent video games exposure and aggression: The role of moral disengagement, anger, hostility, and disinhibition,” *Aggressive behavior*, vol. 45, no. 6, pp. 662–670, 2019.
- [338] N. Maroney, B. J. Williams, A. Thomas, J. Skues, and R. Moulding, “A stress-coping model of problem online video game use,” *International Journal of Mental Health and Addiction*, vol. 17, no. 4, pp. 845–858, 2019.
- [339] N. Johannes, M. Vuorre, and A. K. Przybylski, “Video game play is positively correlated with well-being,” *PsyArXiv. November*, vol. 13, 2020.
- [340] L. Zhu, “The psychology behind video games during COVID-19 pandemic: A case study of Animal Crossing: New Horizons,” *Human Behavior and Emerging Technologies*, 2020.
- [341] E. Ben Simon, A. Rossi, A. G. Harvey, and M. P. Walker, “Overanxious and under-slept,” *Nature human behaviour*, vol. 4, no. 1, pp. 100–110, 2020.
- [342] C. McCall, G. Schofield, D. Halgarth, G. Blyth, A. Laycock, and D. J. Palombo, “The underwood project: A virtual environment for eliciting ambiguous threat,” *Behavior research methods*, pp. 1–16, 2022.
- [343] E. C. Sullivan, C. McCall, M. Croissant, L.-M. Henderson, G. Schofield, and S. A. Cairney, *Tired and Wired: Sleep Deprivation Prevents Affective Renormalization during Exposure to Ambiguous Threat*, Unpublished Manuscript, 2024.
- [344] C. D. Spielberger, “State-trait anxiety inventory for adults,” 1983.
- [345] G. N. Pires, A. G. Bezerra, S. Tufik, and M. L. Andersen, “Effects of acute sleep deprivation on state anxiety levels: a systematic review and meta-analysis,” *Sleep medicine*, vol. 24, pp. 109–118, 2016.
- [346] E. C. Sullivan, “Investigating the cognitive mechanisms by which sleep supports emotion regulation and mental health,” Ph.D. dissertation, University of York, 2023.
- [347] R. Roberts, “Fear of the unknown: Music and sound design in psychological horror games,” in *Music In Video Games*, Routledge, 2014, pp. 138–150.

- [348] T. C. Demarque and E. S. Lima, “Auditory hallucination: Audiological perspective for horror games,” *SBC-Proceedings of SBGames*, vol. 2013, 2013.
- [349] E. Kirkland, “Survival Horrality: Analysis of a Videogame Genre (1),” *Irish Gothic Journal*, no. 10, p. 22, 2011.
- [350] K. F. Steinmetz, “Carceral horror: Punishment and control in Silent Hill,” *Crime, Media, Culture*, vol. 14, no. 2, pp. 265–287, 2018.
- [351] J. M. Girard and A. G. C. Wright, “DARMA: Software for dual axis rating and media annotation,” *Behavior research methods*, vol. 50, pp. 902–909, 2018.
- [352] A. T. Beck, R. A. Steer, G. K. Brown, *et al.*, *Beck depression inventory*. Harcourt Brace Jovanovich New York: 1987.
- [353] A. T. Beck, N. Epstein, G. Brown, and R. Steer, “Beck anxiety inventory,” *Journal of consulting and clinical psychology*, 1993.
- [354] J. A. Horne and O. Ostberg, “A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms,” *International journal of chronobiology*, vol. 4, no. 2, pp. 97–110, 1976.
- [355] D. J. Buysse, C. F. Reynolds III, T. H. Monk, *et al.*, “The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research,” *Psychiatry research*, vol. 28, no. 2, pp. 193–213, 1989. DOI: 10.1007/s12671-010-0033-2.
- [356] D. D. Suhr, “Principal component analysis vs. exploratory factor analysis,” *SUGI 30 proceedings*, vol. 203, no. 230, pp. 1–11, 2005.
- [357] C. Labrín and F. Urdinez, “Principal component analysis,” in *R for political data science*, Chapman and Hall/CRC, 2020, pp. 375–393.
- [358] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in Python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [359] R. Alfarra, A. I. Fins, I. Chayo, and J. L. Tartar, “Changes in attention to an emotional task after sleep deprivation: neurophysiological and behavioral findings,” *Biological psychology*, vol. 104, pp. 1–7, 2015.
- [360] M. Binz and E. Schulz, “Using cognitive psychology to understand GPT-3,” *Proceedings of the National Academy of Sciences*, vol. 120, no. 6, e2218523120, 2023.
- [361] M. Kosinski, “Theory of mind may have spontaneously emerged in large language models,” *arXiv preprint arXiv:2302.02083*, 2023.

- [362] P. Hämmäläinen, M. Tavast, and A. Kunnari, “Evaluating large language models in generating synthetic hci research data: a case study,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–19.
- [363] K. A. Lindquist, A. B. Satpute, and M. Gendron, “Does language do more than communicate emotion?” *Current directions in psychological science*, vol. 24, no. 2, pp. 99–108, 2015.
- [364] J. A. Brooks, H. Shablack, M. Gendron, A. B. Satpute, M. H. Parrish, and K. A. Lindquist, “The role of language in the experience and perception of emotion: A neuroimaging meta-analysis,” *Social Cognitive and Affective Neuroscience*, vol. 12, no. 2, pp. 169–183, 2017.
- [365] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge university press, 1990.
- [366] T. Bosse and E. Zwanenburg, “There’s always hope: Enhancing agent believability through expectation-based emotions,” in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, IEEE, 2009, pp. 1–8.
- [367] K. R. Scherer, A. Schorr, and T. Johnstone, *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press, 2001.
- [368] J. J. Gross and O. P. John, “Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being.,” *Journal of personality and social psychology*, vol. 85, no. 2, p. 348, 2003.
- [369] T. Brown, B. Mann, N. Ryder, *et al.*, “Language models are few-shot learners,” *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [370] R. Bommasani, D. A. Hudson, E. Adeli, *et al.*, “On the opportunities and risks of foundation models,” *arXiv preprint arXiv:2108.07258*, 2021.
- [371] J. Freiknecht and W. Effelsberg, “Procedural generation of interactive stories using language models,” in *Proceedings of the 15th International Conference on the Foundations of Digital Games*, 2020, pp. 1–8.
- [372] S. Yao, R. Rao, M. Hausknecht, and K. Narasimhan, “Keep calm and explore: Language models for action generation in text-based games,” *arXiv preprint arXiv:2010.02903*, 2020.

-
- [373] C. Callison-Burch, G. S. Tomar, L. J. Martin, D. Ippolito, S. Bailis, and D. Reitter, “Dungeons and dragons as a dialog challenge for artificial intelligence,” *arXiv preprint arXiv:2210.07109*, 2022.
- [374] J. S. Park, J. C. O’Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, “Generative agents: Interactive simulacra of human behavior,” *arXiv preprint arXiv:2304.03442*, 2023.
- [375] E. Arts, *The Sims*, [Game], 2000.
- [376] M. Ciolino, J. Kalin, and D. Noever, “The go transformer: natural language modeling for game play,” in *2020 Third International Conference on Artificial Intelligence for Industries (AI4I)*, IEEE, 2020, pp. 23–26.
- [377] T. Sun, J. He, X. Qiu, and X. Huang, “BERTScore is unfair: On social bias in language model-based metrics for text generation,” *arXiv preprint arXiv:2210.07626*, 2022.
- [378] J. Von Oswald, E. Niklasson, E. Randazzo, *et al.*, “Transformers learn in-context by gradient descent,” in *International Conference on Machine Learning*, PMLR, 2023, pp. 35 151–35 174.
- [379] J. Wei, X. Wang, D. Schuurmans, *et al.*, “Chain-of-thought prompting elicits reasoning in large language models,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 824–24 837, 2022.
- [380] M. Singh, V. SB, N. Malviya, *et al.*, “Mind meets machine: Unravelling GPT-4’s cognitive psychology,” *arXiv preprint arXiv:2303.11436*, 2023.
- [381] J. White, Q. Fu, S. Hays, *et al.*, “A prompt pattern catalog to enhance prompt engineering with chatgpt,” *arXiv preprint arXiv:2302.11382*, 2023.
- [382] P. Salovey and J. D. Mayer, “Emotional intelligence,” *Imagination, cognition and personality*, vol. 9, no. 3, pp. 185–211, 1990.
- [383] S. Newsome, A. L. Day, and V. M. Catano, “Assessing the predictive validity of emotional intelligence,” *Personality and Individual Differences*, vol. 29, no. 6, pp. 1005–1016, 2000.

- [384] P. K. Papadogiannis, D. Logan, and G. Sitarenios, “An ability model of emotional intelligence: A rationale, description, and application of the Mayer Salovey Caruso Emotional Intelligence Test (MSCEIT),” *Assessing emotional intelligence: Theory, research, and applications*, pp. 43–65, 2009.
- [385] J. M. Conte, “A review and critique of emotional intelligence measures,” *Journal of organizational behavior*, vol. 26, no. 4, pp. 433–440, 2005.
- [386] C. MacCann and R. D. Roberts, “New paradigms for assessing emotional intelligence: theory and data,” *Emotion*, vol. 8, no. 4, p. 540, 2008.
- [387] P. Sajjadi, L. Hoffmann, P. Cimiano, and S. Kopp, “A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users,” *Entertainment Computing*, vol. 32, p. 100 313, 2019.
- [388] *OpenAI API*. [Online]. Available: <https://platform.openai.com/overview>.
- [389] J. W. Pennebaker, M. E. Francis, and R. J. Booth, “Linguistic inquiry and word count: LIWC 2001,” *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.
- [390] J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson, “Measuring emotional expression with the Linguistic Inquiry and Word Count,” *The American journal of psychology*, vol. 120, no. 2, pp. 263–286, 2007.
- [391] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, “Lying words: Predicting deception from linguistic styles,” *Personality and social psychology bulletin*, vol. 29, no. 5, pp. 665–675, 2003.
- [392] M. A. Cohn, M. R. Mehl, and J. W. Pennebaker, “Linguistic markers of psychological change surrounding September 11, 2001,” *Psychological science*, vol. 15, no. 10, pp. 687–693, 2004.
- [393] K. Krippendorff, *Content analysis: An introduction to its methodology*. Sage publications, 2018.
- [394] *OpenAI Moderation API*. [Online]. Available: <https://platform.openai.com/docs/guides/moderation>.
- [395] H. A. Elfenbein, S. G. Barsade, and N. Eisenkraft, “The social perception of emotional abilities: expanding what we know about observer ratings of emotional intelligence,” *Emotion*, vol. 15, no. 1, p. 17, 2015.

- [396] K. S. Law, C.-S. Wong, and L. J. Song, “The construct and criterion validity of emotional intelligence and its potential utility for management studies,” *Journal of applied Psychology*, vol. 89, no. 3, p. 483, 2004.
- [397] A. J. C. Cuddy, S. T. Fiske, and P. Glick, “The BIAS map: behaviors from intergroup affect and stereotypes,” *Journal of personality and social psychology*, vol. 92, no. 4, p. 631, 2007.
- [398] M. Croissant, *No Title*.
- [399] M. C. Frank, “Large language models as models of human cognition,” *PsyArXiv*, 2023.
- [400] C. Li, J. Wang, K. Zhu, *et al.*, “Emotionprompt: Leveraging psychology for large language models enhancement via emotional stimulus,” *arXiv preprint arXiv:2307.11760*, 2023.
- [401] S. A. Kotz and S. Paulmann, “Emotion, language, and the brain,” *Language and Linguistics Compass*, vol. 5, no. 3, pp. 108–125, 2011.
- [402] K. Isbister, *Better game characters by design: A psychological approach*. CRC Press, 2022.
- [403] A. Sobieszek and T. Price, “Playing games with AIs: the limits of GPT-3 and similar large language models,” *Minds and Machines*, vol. 32, no. 2, pp. 341–364, 2022.
- [404] T. Wu, M. Terry, and C. J. Cai, “Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts,” in *Proceedings of the 2022 CHI conference on human factors in computing systems*, 2022, pp. 1–22.
- [405] O. Topsakal and T. C. Akinci, “Creating large language model applications utilizing langchain: A primer on developing llm apps fast,” in *Proceedings of the International Conference on Applied Engineering and Natural Sciences, Konya, Turkey*, 2023, pp. 10–12.
- [406] Y. Han, C. Liu, and P. Wang, “A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge,” *arXiv preprint arXiv:2310.11703*, 2023.
- [407] H. Liu, R. Ning, Z. Teng, J. Liu, Q. Zhou, and Y. Zhang, “Evaluating the logical reasoning ability of chatgpt and gpt-4,” *arXiv preprint arXiv:2304.03439*, 2023.

- [408] E. Hudlicka, “To feel or not to feel: The role of affect in human–computer interaction,” *International journal of human-computer studies*, vol. 59, no. 1-2, pp. 1–32, 2003.
- [409] A. Wierzbicka, “Talking about emotions: Semantics, culture, and cognition,” *Cognition & Emotion*, vol. 6, no. 3-4, pp. 285–319, 1992.
- [410] D. Charles, M. McNeill, M. McAlister, *et al.*, “Player-centred game design: Player modelling and adaptive digital games,” 2005.
- [411] S. Tognetti, M. Garbarino, A. Bonarini, and M. Matteucci, “Modeling enjoyment preference from physiological responses in a car racing game,” in *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games*, 2010, pp. 321–328.
- [412] H. P. Martínez and G. N. Yannakakis, “Mining multimodal sequential patterns: A case study on affect detection,” in *ICMI’11 - Proceedings of the 2011 ACM International Conference on Multimodal Interaction*, 2011, pp. 3–10. DOI: 10.1145/2070481.2070485.
- [413] P. Honey, A. Mumford, *et al.*, *The manual of learning styles*. Peter Honey Maidenhead, 1992, vol. 3.
- [414] A. Aleksieva-Petrova, M. Petrov, and B. Bontchev, “Game and learner ontology model,” in *Int. Scientific Conf. Computer Science’2011*, 2011, pp. 1–2.
- [415] B. Yuan, E. Folmer, and F. C. Harris, “Game accessibility: a survey,” *Universal Access in the information Society*, vol. 10, no. 1, pp. 81–100, 2011.